

**IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF DELAWARE**

AFFYMETRIX, INC., a Delaware
corporation,

Plaintiff,

v.

ILLUMINA, INC., a Delaware corporation,
Defendant.

Civil Case No.

JURY TRIAL DEMANDED

COMPLAINT FOR PATENT INFRINGEMENT

Plaintiff AFFYMETRIX, INC. ("Affymetrix"), for its complaint against defendant ILLUMINA, INC. ("Illumina"), alleges as follows:

NATURE OF THE ACTION

1. This is an action under the patent laws of the United States, 35 U.S.C. §§ 1, *et seq.*, for infringement by Illumina of patents owned by Affymetrix.

THE PARTIES

2. Plaintiff Affymetrix is a Delaware corporation with its principal place of business at 3420 Central Expressway, Santa Clara, California 95051.

3. On information and belief, Defendant Illumina is a corporation incorporated under the laws of the state of Delaware with its principal place of business at 9885 Towne Centre Drive, San Diego, California 92121. On information and belief, Illumina offers for sale and/or sells its infringing products in this Judicial District, among other places.

JURISDICTION AND VENUE

4. This Court has jurisdiction over the subject matter of this action pursuant to 28 U.S.C. §§ 1331 and 1338(a).

5. This Court has personal jurisdiction over Illumina because Illumina is incorporated in the State of Delaware and/or has purposely availed itself of the privilege of conducting activities within this State and District.

6. Venue is proper in this judicial district pursuant to 28 U.S.C. §§ 1391(b) & (c) and 1400(b).

BACKGROUND AND THE PATENTS-IN-SUIT

7. Dr. Stephen Fodor and his team of scientists spearheaded the development of DNA microarray technology at Affymax, N.V. (“Affymax”) in the late 1980s. This technology became the basis for a new company, Affymetrix, formed as a division of Affymax, in 1991. Affymetrix began operating independently in 1992.

8. Affymetrix is a pioneer in the DNA microarray industry and develops state-of-the-art technology for acquiring, analyzing, and managing complex genetic information for use in biomedical research, genomics, and clinical diagnostics. Affymetrix’s GeneChip® microarrays are the leading commercial application in this field. GeneChip® microarrays consist of many known sequences of DNA attached to a substrate that hybridize to complementary genetic sequences in experimental samples. Researchers use the GeneChip® microarrays to develop new therapeutic drugs, investigate the cause of cancer, AIDS, and other life-threatening diseases, and explore the human genome.

9. In addition to the GeneChip® microarray, Affymetrix has also developed complementary instruments and software for analyzing nucleic acids. These instruments include the GeneChip® Fluidics Station, which is used to wash away unhybridized sample DNA from the GeneChip® microarray and stain the hybridized DNA with a fluorescent molecule, and the GeneChip® Scanner, which detects where on the array the fluorescent signal is present.

Software developed by Affymetrix includes the GeneChip® Analysis Suite, which is used to collect and analyze data generated by the GeneChip® microarrays.

10. Affymetrix made substantial investments in research and development in order to pioneer the commercial market in DNA microarrays and to achieve a position of technological leadership. Affymetrix owns valuable intellectual property rights in DNA microarray and related technology.

11. Among these intellectual property rights are the following United States patents:

- A) U.S. Patent No. 5,902,723 (the “‘723 patent”), issued on May 11, 1999, entitled “Analysis of surface immobilized polymers utilizing microfluorescence detection.” A true and correct copy of the ‘723 patent is attached to this Complaint as Exhibit 1 and is incorporated herein by reference.
- B) U.S. Patent No. 6,403,320 (the “‘320 patent”), issued on June 11, 2002, entitled “Support bound probes and methods of analysis using the same.” A true and correct copy of the ‘320 patent is attached to this Complaint as Exhibit 2 and is incorporated herein by reference.
- C) U.S. Patent No. 6,420,169 (the “‘169 patent”), issued on July 16, 2002, entitled “Apparatus for forming polynucleotides or polypeptides.” A true and correct copy of the ‘169 patent is attached to this Complaint as Exhibit 3 and is incorporated herein by reference.
- D) U.S. Patent No. 6,576,424 (the “‘424 patent”), issued on June 10, 2003, entitled “Arrays and methods for detecting nucleic acids.” A true and

correct copy of the '424 patent is attached to this Complaint as Exhibit 4 and is incorporated herein by reference.

- E) U.S. Patent No. 7,056,666 (the "'666 patent"), issued on June 6, 2006, entitled "Analysis of surface immobilized polymers utilizing microfluorescence detection." A true and correct copy of the '666 patent is attached to this Complaint as Exhibit 5 and is incorporated herein by reference.

12. The '723, '320, '169, '424, and '666 patents were duly issued and are owned by Affymetrix (the '723, '320, '169, '424, and '666 patents are hereinafter collectively referred to as the "Patents-in-Suit"). Affymetrix has the full legal right to sue, enforce, and recover damages for all infringements of the Patents-in-Suit.

HISTORY BETWEEN THE PARTIES

13. After learning about the exciting DNA microarray technology being developed at Affymetrix, Dr. Mark Chee, who had recently completed his graduate studies, approached Dr. Stephen Fodor, Affymetrix's founder and current Chief Executive Officer, about doing post-doctorate work at Affymetrix. Dr. Chee had never worked with DNA microarray technology prior to his association with Affymetrix. Dr. Fodor agreed to allow Dr. Chee to do his post-doctorate work on DNA microarray technology at Affymetrix.

14. Shortly thereafter, Dr. Chee became an employee at Affymetrix. He worked at Affymetrix from 1993 until 1997, becoming an inventor on several Affymetrix patents. Dr. Chee rose to the position of Director of Genetics Research at Affymetrix. During this time, Dr. Chee supervised many other scientists at Affymetrix, including Drs. Kevin Gunderson and Jian-Bing Fan. While at Affymetrix, Dr. Fan developed a tag-based genotyping assay and ultimately

published a paper describing that assay. Dr. Gunderson worked on extension- and ligation-based genotyping assays at Affymetrix. Neither Dr. Gunderson nor Dr. Fan had experience working with DNA microarray technology prior to arriving at Affymetrix.

15. In 1997, Dr. Chee ended his employment at Affymetrix. Shortly thereafter, Dr. Chee co-founded Illumina. Dr. Chee played a significant role in developing Illumina's DNA microarray technology. He also was instrumental in attracting several other Affymetrix scientists to join Illumina, including Drs. Gunderson and Fan. Drs. Gunderson and Fan, in turn, were the key architects of Illumina's DNA microarray genotyping assays. Specifically, Dr. Fan developed Illumina's tag-based genotyping GoldenGate assay while Dr. Gunderson developed Illumina's extension-based genotyping Infinium I and II assays. Today, Illumina makes and sells DNA microarrays, sequencing applications, and related technology that directly compete with Affymetrix's products.

16. During the 1990s, Affymetrix collaborated with a company called Molecular Dynamics, Inc ("MD"). Jay Flatley was the CEO of MD. By 1997, MD became interested in manufacturing and selling DNA microarrays. To that end, MD requested and then took a license to Affymetrix's DNA microarray patent portfolio. Mr. Flatley signed that license on behalf of MD. After leaving MD, Mr. Flatley became the CEO of Illumina.

17. When Affymetrix learned in 2002 that Illumina was planning to market DNA microarray technology and related products, Affymetrix approached Illumina about taking a license to Affymetrix's patents. Despite reviewing many Affymetrix patents, Illumina chose to disregard them and continue developing its technology without a license.

18. In March 2004, Affymetrix again approached Illumina about taking a license to Affymetrix's patent portfolio. During a series of meetings on this issue, Affymetrix specifically

identified many Affymetrix patents to which Illumina needed a license. These patents included the '320 and '424 patents. Illumina chose to disregard these patents and continue developing, marketing, and selling its products without a license.

19. Left with no alternative, Affymetrix filed a patent infringement suit against Illumina on July 26, 2004, in the United States District Court for the District of Delaware. On March 13, 2007, a jury returned a verdict that Illumina infringed each of the asserted claims of the five patents. The jury also imposed a 15% royalty rate and damages through 2005 of \$16,727,459. Notwithstanding the verdict, Illumina continued to make and sell its infringing products without a license from Affymetrix.

20. In November 2006, Illumina announced its intention to purchase Solexa, Inc. ("Solexa"), a genetic sequencing company. On December 5, 2006, Affymetrix once again approached Illumina about taking a license to Affymetrix's patent portfolio. In the related correspondence, Affymetrix identified many specific patents to which Illumina needed a license, including the '723, '320, '169, '424, and '666 patents. Affymetrix also notified Illumina that many of these patents, including the '169 and '666 patents, were relevant to the Solexa technology that Illumina planned to acquire. After initially agreeing to a meeting at which these issues would be discussed, Illumina canceled the meeting and continued to develop, market, and sell its products without a license.

21. After receiving no further response from Illumina, on December 26, 2006, Affymetrix approached Solexa directly about taking a license to certain Affymetrix patents. These patents included the '169, '424, and '666 patents. Solexa responded on January 4, 2007, by indicating that it was reviewing the referenced patents and that future discussion on the issue would involve Illumina, which was in the process of completing its acquisition of Solexa.

Illumina completed its acquisition of Solexa in January 2007. Neither Solexa nor Illumina ever followed up on this initial licensing correspondence.

COUNT 1

(Infringement of United States Patent No. 5,902,723 by Illumina)

22. Affymetrix realleges and incorporates herein by reference the allegations stated in paragraphs 1-21 of this Complaint.

23. Illumina has been and still is infringing the '723 patent by making, using, offering for sale, and/or selling assays, products, software, and associated instrumentation under the name BeadArray™ technology, including, but not limited to, the Infinium® II assay.

24. Illumina also has been and is contributorily infringing and/or actively inducing others to infringe the '723 patent by supplying the aforementioned BeadArray technology and associated instrumentation.

25. On information and belief, Illumina's infringement, contributory infringement, and/or active inducement of others' infringement of the '723 patent has taken place with full knowledge of the '723 patent and has been intentional, deliberate, and willful.

26. On information and belief, Illumina will continue to infringe, contributorily infringe, and/or actively induce others to infringe the '723 patent unless and until it is enjoined by this Court.

COUNT 2

(Infringement of United States Patent No. 6,403,320 by Illumina)

27. Affymetrix realleges and incorporates herein by reference the allegations stated in paragraphs 1-21 of this Complaint.

28. Illumina has been and still is infringing the '320 patent by making, using, offering for sale, and/or selling products and associated instrumentation under the name BeadArray™ technology, including, but not limited to, all of Illumina's Sentrix® Array Matrix and BeadChip arrays, the BeadArray Reader, and the detection instrument used to decode all of Illumina's BeadArray arrays.

29. Illumina also has been and is contributorily infringing and/or actively inducing others to infringe the '320 patent by supplying the BeadArray technology and associated instrumentation.

30. On information and belief, Illumina's infringement, contributory infringement, and/or active inducement of others' infringement of the '320 patent has taken place with full knowledge of the '320 patent and has been intentional, deliberate, and willful.

31. On information and belief, Illumina will continue to infringe, contributorily infringe, and/or actively induce others to infringe the '320 patent unless and until it is enjoined by this Court.

COUNT 3

(Infringement of United States Patent No. 6,420,169 by Illumina)

32. Affymetrix realleges and incorporates herein by reference the allegations stated in paragraphs 1-21 of this Complaint.

33. Illumina has been and still is infringing the '169 patent by making, using, offering for sale, and/or selling products and associated software and instrumentation under the name Solexa Sequencing technology, including, but not limited to, the Genome Analyzer, Clonal Single Molecule Array technology, and the 1G Genome Analysis System.

34. Illumina also has been and is contributorily infringing and/or actively inducing others to infringe the '169 patent by supplying the aforementioned Solexa Sequencing technology.

35. On information and belief, Illumina's infringement, contributory infringement, and/or active inducement of others' infringement of the '169 patent has taken place with full knowledge of the '169 patent and has been intentional, deliberate, and willful.

36. On information and belief, Illumina will continue to infringe, contributorily infringe, and/or actively induce others to infringe the '169 patent unless and until it is enjoined by this Court.

COUNT 4

(Infringement of United States Patent No. 6,576,424 by Illumina)

37. Affymetrix realleges and incorporates herein by reference the allegations stated in paragraphs 1-21 of this Complaint.

38. Illumina has been and still is infringing the '424 patent by making, using, offering for sale, and/or selling products and associated instrumentation under the name BeadArray™ technology, including, but not limited to, BeadArray arrays used with the Infinium I and II and Direct Hyb assays.

39. Illumina also has been and is contributorily infringing and/or actively inducing others to infringe the '424 patent by supplying the BeadArray technology and associated instrumentation.

40. On information and belief, Illumina's infringement, contributory infringement, and/or active inducement of others' infringement of the '424 patent has taken place with full knowledge of the '424 patent and has been intentional, deliberate, and willful.

41. On information and belief, Illumina will continue to infringe, contributorily infringe, and/or actively induce others to infringe the '424 patent unless and until it is enjoined by this Court.

COUNT 5

(Infringement of United States Patent No. 7,056,666 by Illumina)

42. Affymetrix realleges and incorporates herein by reference the allegations stated in paragraphs 1-21 of this Complaint.

43. Illumina has been and still is infringing the '666 patent by making, using, offering for sale, and/or selling products and associated software and instrumentation under the name Solexa Sequencing technology, including, but not limited to, the Genome Analyzer, Clonal Single Molecule Array technology, and the 1G Genome Analysis System.

44. Illumina also has been and is contributorily infringing and/or actively inducing others to infringe the '666 patent by supplying the aforementioned Solexa Sequencing technology.

45. On information and belief, Illumina's infringement, contributory infringement, and/or active inducement of others' infringement of the '666 patent has taken place with full knowledge of the '666 patent and has been intentional, deliberate, and willful.

46. On information and belief, Illumina will continue to infringe, contributorily infringe, and/or actively induce others to infringe the '666 patent unless and until it is enjoined by this Court.

PRAYER FOR RELIEF

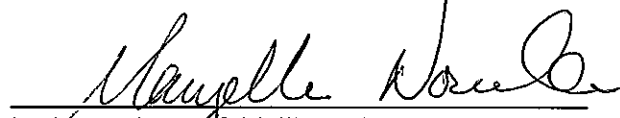
WHEREFORE, Plaintiff Affymetrix prays that this Court:

- A. Enter a judgment that Illumina has infringed, contributorily infringed, and actively induced others to infringe Affymetrix's Patents-in-Suit;
- B. Grant a permanent injunction restraining and enjoining Illumina, its officers, directors, agents, servants, employees, successors, assigns, parent, subsidiaries, affiliated or related companies, and attorneys from infringing, inducing others to infringe and contributing to the infringement of Affymetrix's Patents-in-Suit.
- C. Award Affymetrix damages in an amount sufficient to compensate Affymetrix for Illumina's infringement, contributory infringement and active inducement of others' infringement of Affymetrix's Patents-in-Suit, but not less than a reasonable royalty;
- D. Award prejudgment interest to Affymetrix pursuant to 35 U.S.C. § 284;
- E. Award increased damages, pursuant to 35 U.S.C. § 284, in an amount not less than three times the amount of actual damages awarded to Affymetrix, by reason of Illumina's willful infringement of Affymetrix's Patents-in-Suit;
- F. Declare this case exceptional under 35 U.S.C. § 285 and award Affymetrix its reasonable attorneys' fees, expenses and costs incurred in this action; and
- G. Grant Affymetrix such other and further relief as this Court may deem just and proper.

DEMAND FOR JURY TRIAL

Affymetrix hereby demands a jury trial on all issues appropriately triable by a jury.

MORRIS, NICHOLS, ARSHT & TUNNELL LLP

A handwritten signature in black ink, appearing to read "Maryellen Noreika", is written over a horizontal line.

Jack B. Blumenfeld (#1014)
Maryellen Noreika (#3208)
1201 North Market Street
P.O. Box 1347
Wilmington, DE 19899
(302) 658-9200

Attorneys for Plaintiff, Affymetrix, Inc.

Of Counsel:

Michael J. Malecek
Daniel R. Reed
Peter E. Root
Stephen C. Holmes
AFFYMETRIX, INC.
6550 Vallejo Street
Suite 100
Emeryville, CA 94608
(510) 428-8500

October 24, 2007

EXHIBIT 1

US005902723A

United States Patent [19][11] **Patent Number:** **5,902,723****Dower et al.**[45] **Date of Patent:** **May 11, 1999**[54] **ANALYSIS OF SURFACE IMMOBILIZED
POLYMERS UTILIZING
MICROFLUORESCENCE DETECTION**[76] Inventors: **William J. Dower**, 761 Partridge Ave.,
Menlo Park, Calif. 94025; **Stephen P.
A. Fodor**, 3863 Nathan Way, Palo Alto,
Calif. 94303[21] Appl. No.: **08/679,478**[22] Filed: **Jul. 12, 1996****Related U.S. Application Data**[63] Continuation of application No. 07/626,730, Dec. 6, 1990,
Pat. No. 5,547,839, which is a continuation-in-part of appli-
cation No. 07/492,462, Mar. 7, 1990, Pat. No. 5,143,854,
which is a continuation-in-part of application No. 07/362,
901, Jun. 7, 1989, abandoned.[51] **Int. Cl.⁶** **C12Q 1/68**[52] **U.S. Cl.** **435/6; 536/23.1; 536/243;
435/188.5**[58] **Field of Search** **435/6, 188.5; 536/23.1,
536/24.3**[56] **References Cited****U.S. PATENT DOCUMENTS**

4,542,102	9/1985	Dattagupta et al.	435/6
4,582,789	4/1986	Sheldon et al.	435/6
4,646,127	2/1987	Mundy	435/6
4,689,405	8/1987	Frank et al.	536/27
4,713,326	12/1987	Dattagupta et al.	435/6
4,855,225	8/1989	Fung et al.	435/6
4,889,818	12/1989	Gelfand et al.	435/194
4,962,037	10/1990	Jett et al.	
4,965,188	10/1990	Mullis et al.	435/6
5,002,867	3/1991	Macevitz	
5,026,840	6/1991	Dattagupta et al.	536/27
5,075,216	12/1991	Innis et al.	435/6
5,126,239	6/1992	Livak et al.	
5,143,854	9/1992	Pirrung et al.	

FOREIGN PATENT DOCUMENTS

392 546	10/1990	European Pat. Off.	
2233654	1/1991	United Kingdom	
WO 89/10977	11/1989	WIPO	
WO 89/11548	11/1989	WIPO	
WO 90/03382	4/1990	WIPO	
WO 90/13666	5/1990	WIPO	C12Q 1/68
WO 90/15070	12/1990	WIPO	
WO 91/06678	5/1991	WIPO	C12Q 1/68
WO 91/07087	5/1991	WIPO	

OTHER PUBLICATIONS

Amit et al., *J. Org. Chem.*, 39(2):192-196 (1974).
 Bains et al., *J. Theor. Biol.*, 135:303-307 (1988).
 Barinaga, *Science*, 253:1489 (1991).
 Carrano et al., *Genomics*, 4:129-136 (1989).
 Chatterjee et al., *J. Am. Chem. Soc.*, 112:6397-6399 (1990).
 Chidgeavadze et al., *Nuc. Acids Res.*, 12(3):1671-1686 (1984).
 Chidgeavadze et al., *FEBS Lett.*, 183(2):275-278 (1985).
 Chien et al., *J. Bacteriol.*, 127:1550-1557 (1976).
 Cimino et al., *Ann. Rev. Biochem.*, 54:1151-1193 (1985).
 Coulson et al., *PNAS*, 83:7821-7825 (1986).

Craig et al., *Nucl. Acids Res.*, 18:2653-2660 (1990).
 Dower and Fodor, *Ann. Rep. Med. Chem.*, 26:271-280 (1991).
 Evans et al., *PNAS*, 86:5030-5034 (1989).
 Fodor et al., *Science*, 251:767-773 (1991).
 Frank et al., *Bio/Tech*, 6:1211-1213 (1988).
 Gerard et al., *DNA*, 5(4):271-279 (1986).
 Houts, *J. Virol.*, 29:517-522 (1979).
 Ikehara et al., *Adv. in Carbohydrate Chemistry and Biochemistry*, 36:135-213 (1979).
 Innis et al., *PNAS*, 85:9436-9440 (1988).
 Jacobsen et al., *Eur. J. Biochem.*, 45:624-627 (1974).
 Kambara et al., *Bio/Tech*, 6:816-821 (1988).
 Klenow and Henningsen, *PNAS*, 65(2):168-175 (1970).
 Knight, *Bio/Tech*, 7:1075-1076 (1989).
 Kotewicz et al., *Gene*, 85:249-258 (1985).
 Kutateladze et al., *Molekulyarnaya Biologiya*, 20:267-276 (1986) (abstract only).
 Little, *Nature*, 346:611-612 (1990).
 Maxam and Gilbert, *Meth. Enz.*, 65:499-560 (1980).
 McCray et al., *Ann. Rev. Biophys. Biophys. Chem.*, 18:239-270 (1989).
 Micheils et al., *Cabios*, 3(3):203-210 (1987).
 Nelson et al., *Nucl. Acids Res.*, 17(18):7179-7186 (1989).
 Nossal, *J. Biol. Chem.*, 249(17):5668-5676 (1974).
 Ohtsuka et al., *Nuc. Acids Res.*, 1(10):1351-1357 (1974).
 Olson et al., *PNAS*, 83:7826-7830 (1986).
 Parsons, *Photochem. Photobiol.*, 32:813-821 (1980).
 Patchornik et al., *J. Am. Chem. Soc.*, 92(21):6333-6335 (1970).
 Pfeifer et al., *Science*, 246:810-812 (1989).
 Poustka et al., *CSH Symp. Quant. Biol.*, 51:131-139 (1986).
 Prober et al., *Science*, 238:336-341 (1987).
 Ross et al., *J. Mol. Biol.*, 201:339-351 (1988).
 Ruth et al., *Mol. Pharm.*, 20:415-422 (1981).
 Saiki and Gelfand, *Amplifications*, 1:4-6 (1989).
 Sanger and Coulson, *J. Mol. Biol.*, 94:441-448 (1975).
 Sanger et al., *PNAS*, 74(12):5463-5467 (1977).
 Seed, *Nuc. Acids Res.*, 10(5):1799-1810 (1982).
 Smith et al., *Nucl. Acids Res.*, 13(7):2399-2412 (1985).
 Smith et al., *Nature*, 321:674-679 (1986).
 Song et al., *Photochem. Photobiol.*, 29:1177-1197 (1979).
 Tabor and Richardson, *PNAS*, 84:4767-4771 (1987).
 Tabor and Richardson, *J. Biol. Chem.*, 262:15330-15333 (1987).
 Tsugita et al., *J. Biochem.*, 106:60-65 (1989).
 Wiesenbahn et al., *PNAS*, 75:2703-2707 (1978).
 Wood et al., *PNAS*, 82:1585-1588 (1985).
 Ye and Hong, *Scientia Sinica*, 30(5):503-506 (1987).

Primary Examiner—Scott W. Houtteman
Attorney, Agent, or Firm—Townsend & Townsend & Crew

[57]

ABSTRACT

Means for simultaneous parallel sequence analysis of a large number of biological polymer macromolecules. Apparatus and methods may use fluorescent labels in repetitive chemistry to determine terminal monomers on solid phase immobilized polymers. Reagents which specifically recognize terminal monomers are used to label polymers at defined positions on a solid substrate.

15 Claims, 10 Drawing Sheets

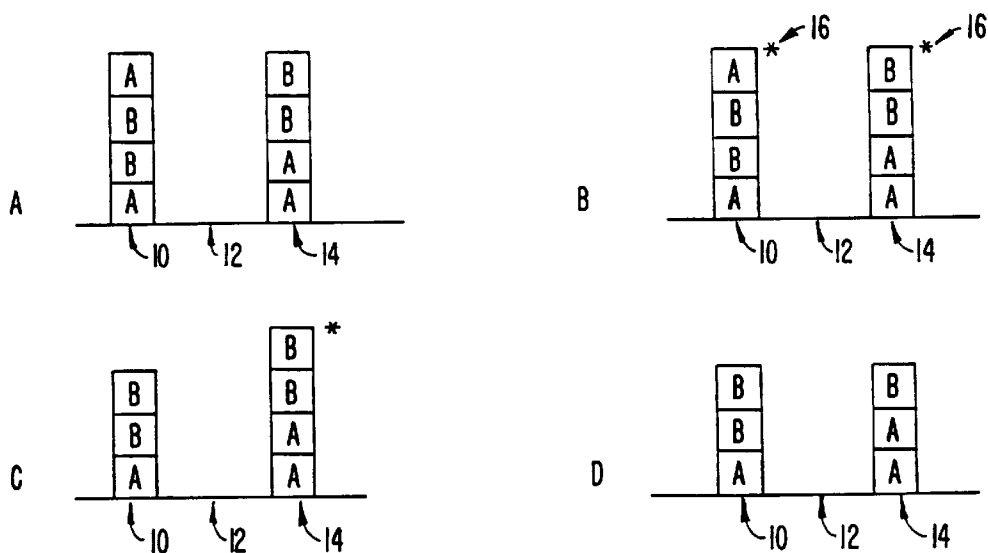


FIG. 1.

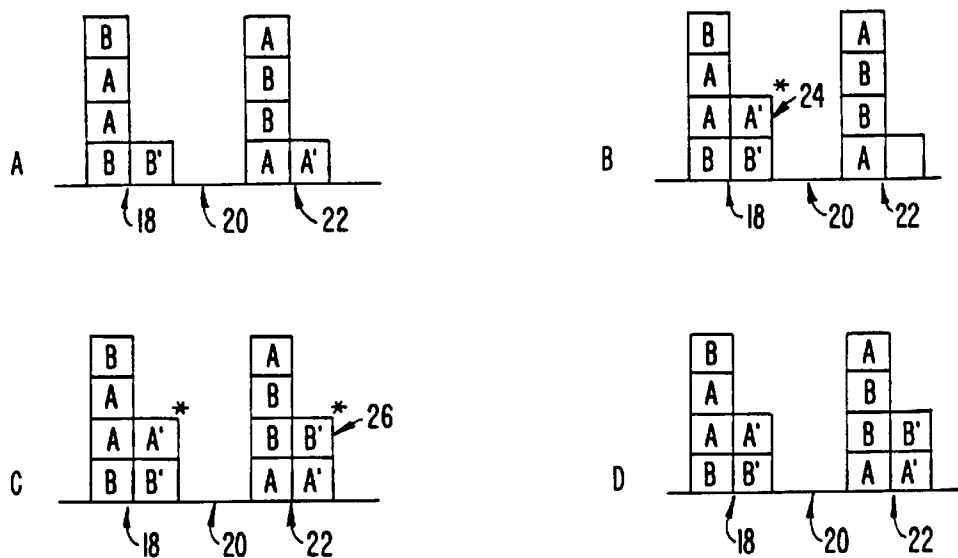


FIG. 2.

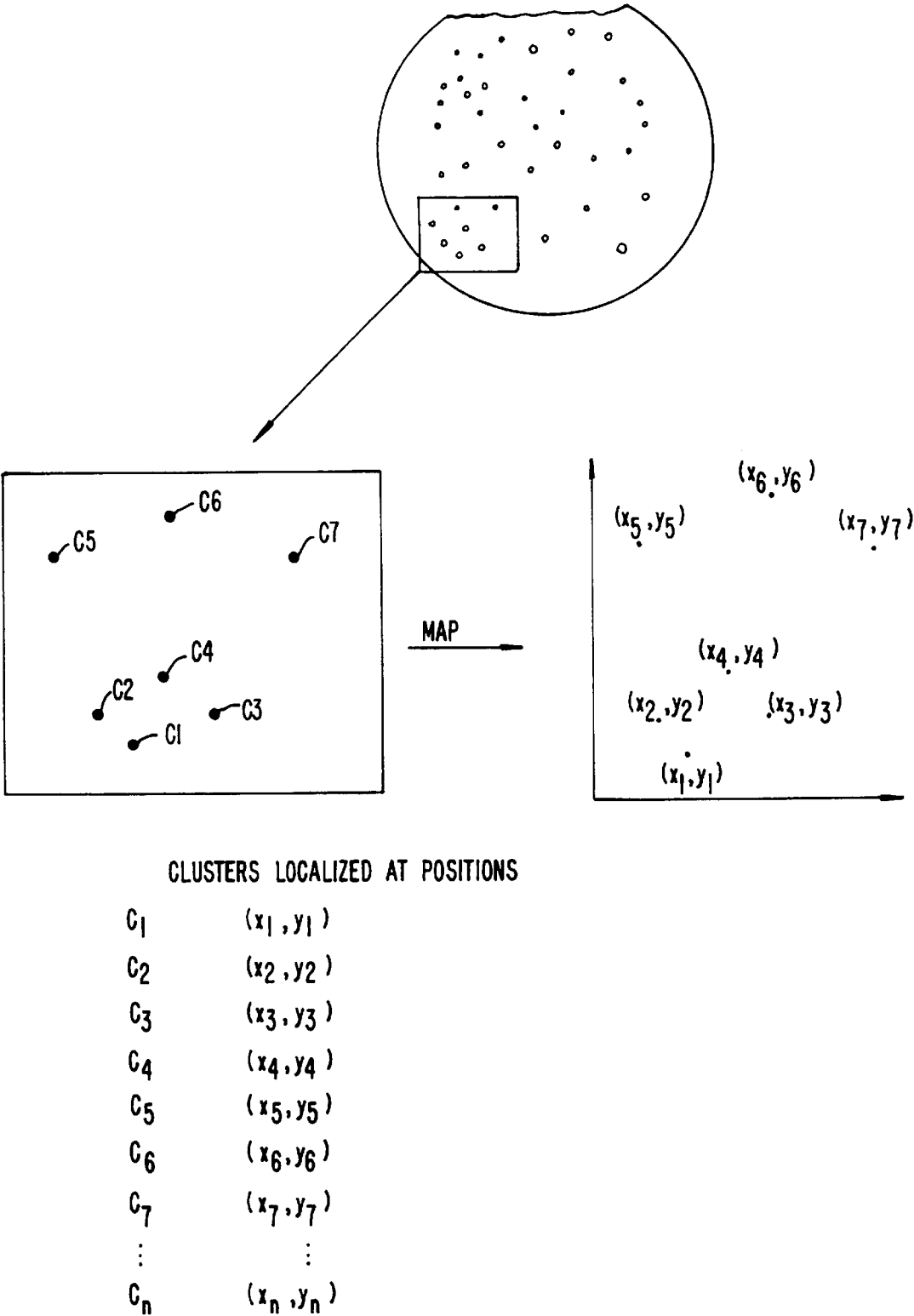


FIG. 3.

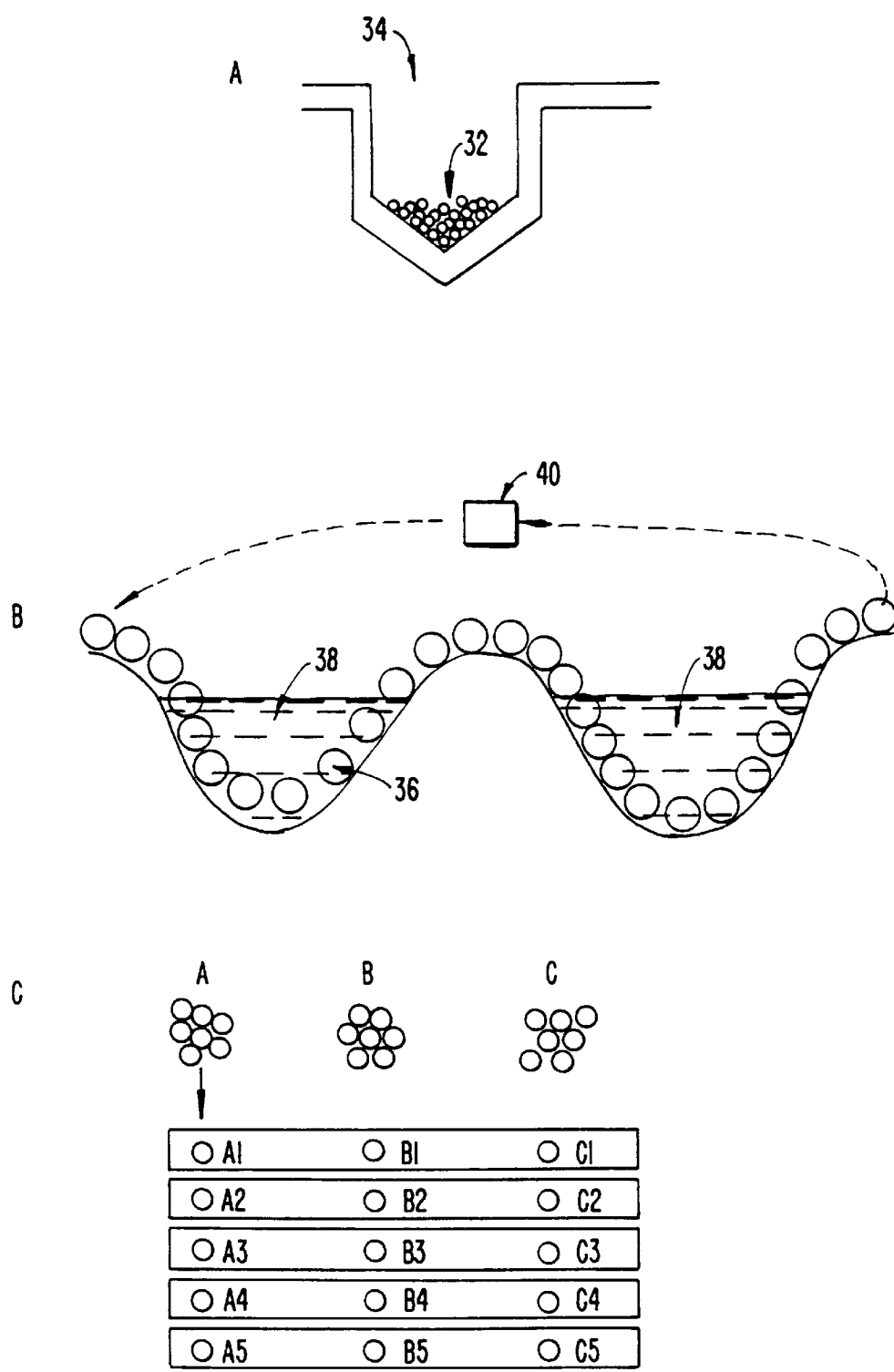


FIG. 4.

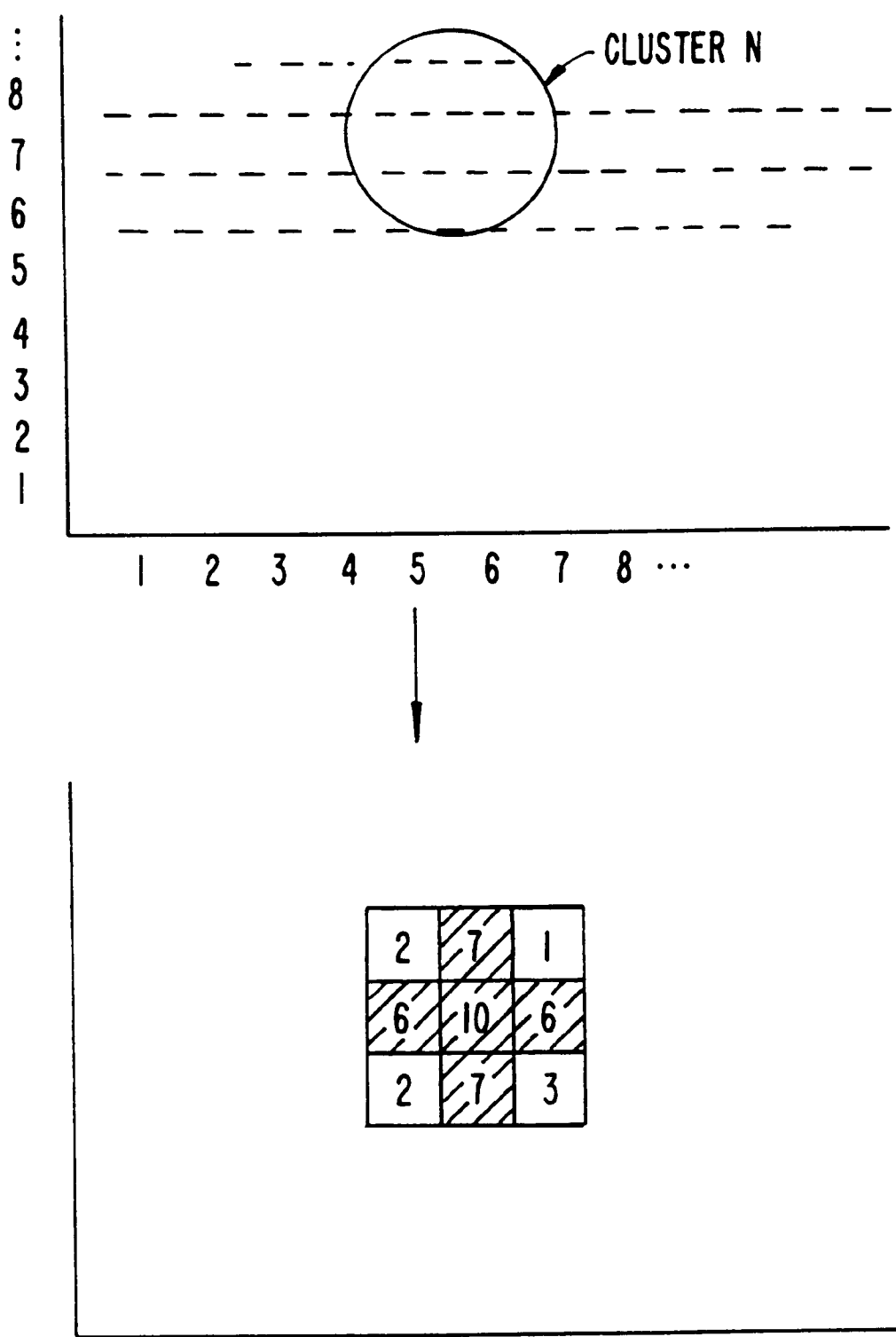
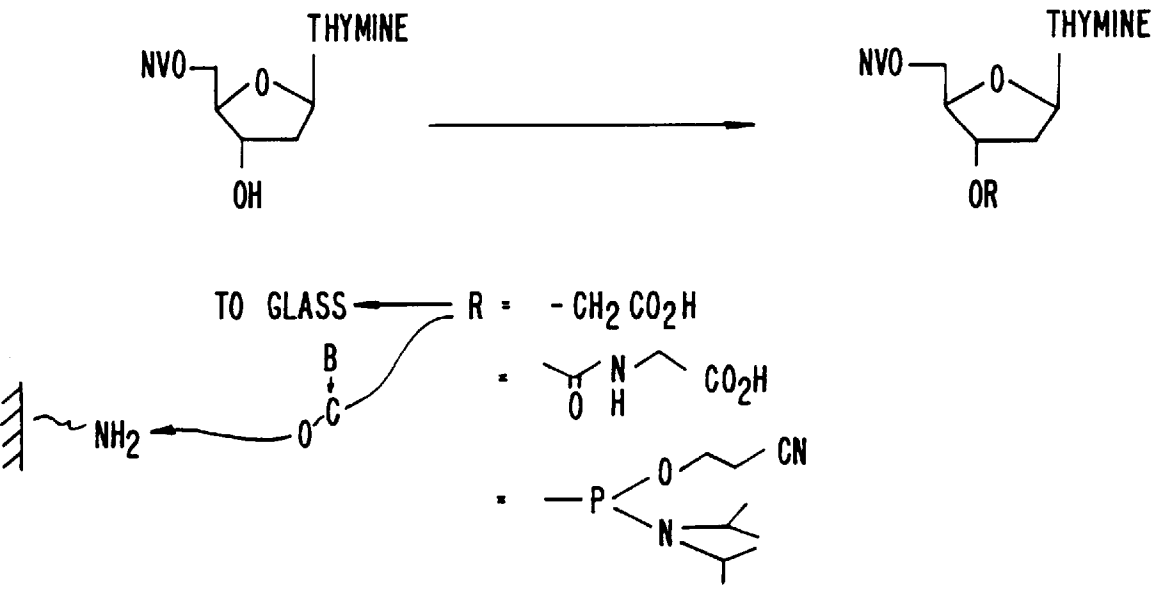


FIG. 5.



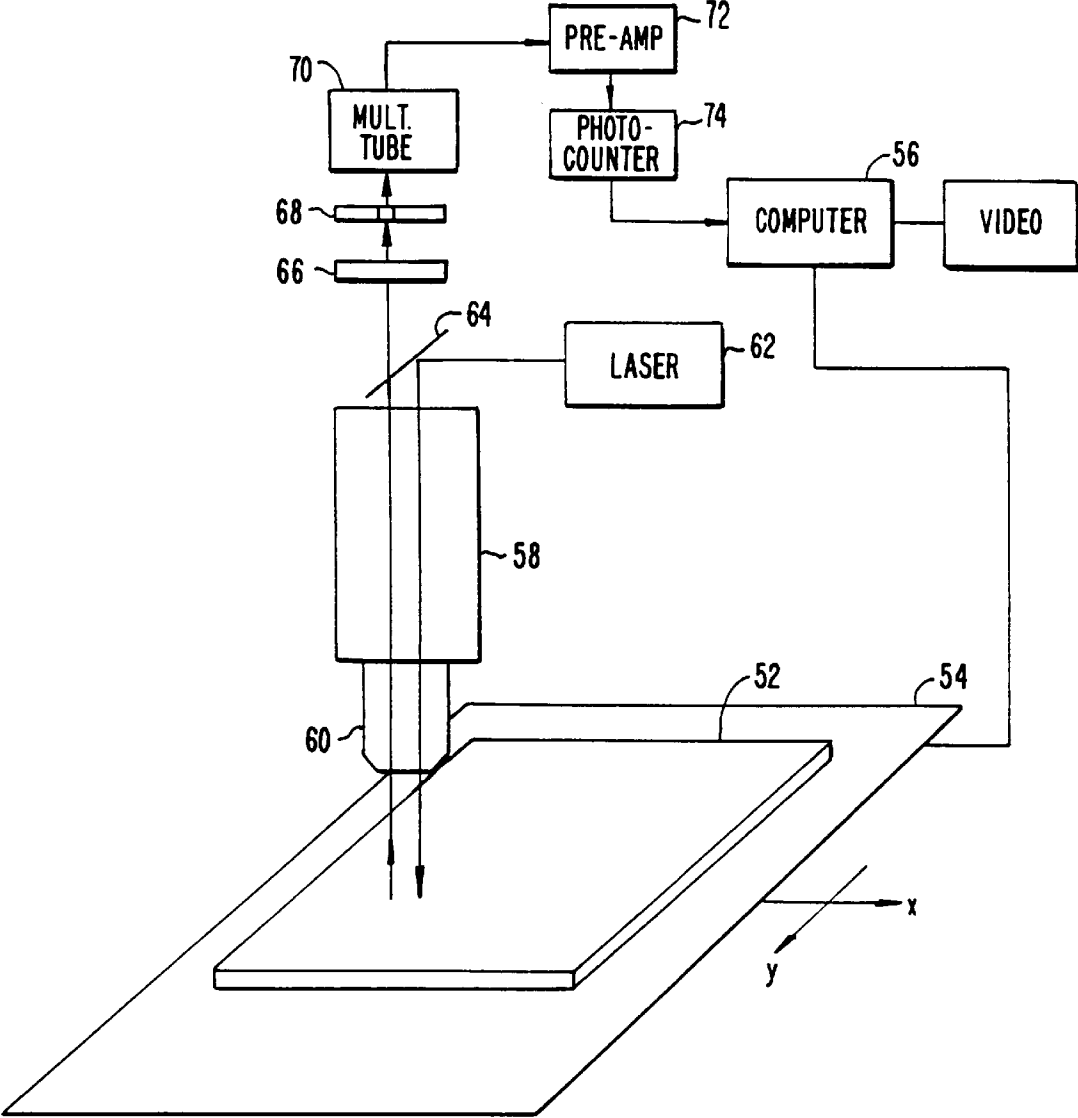


FIG. 7.

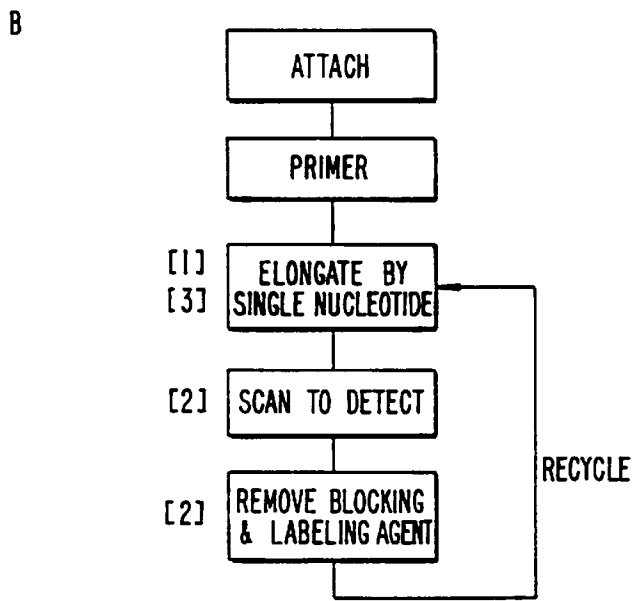
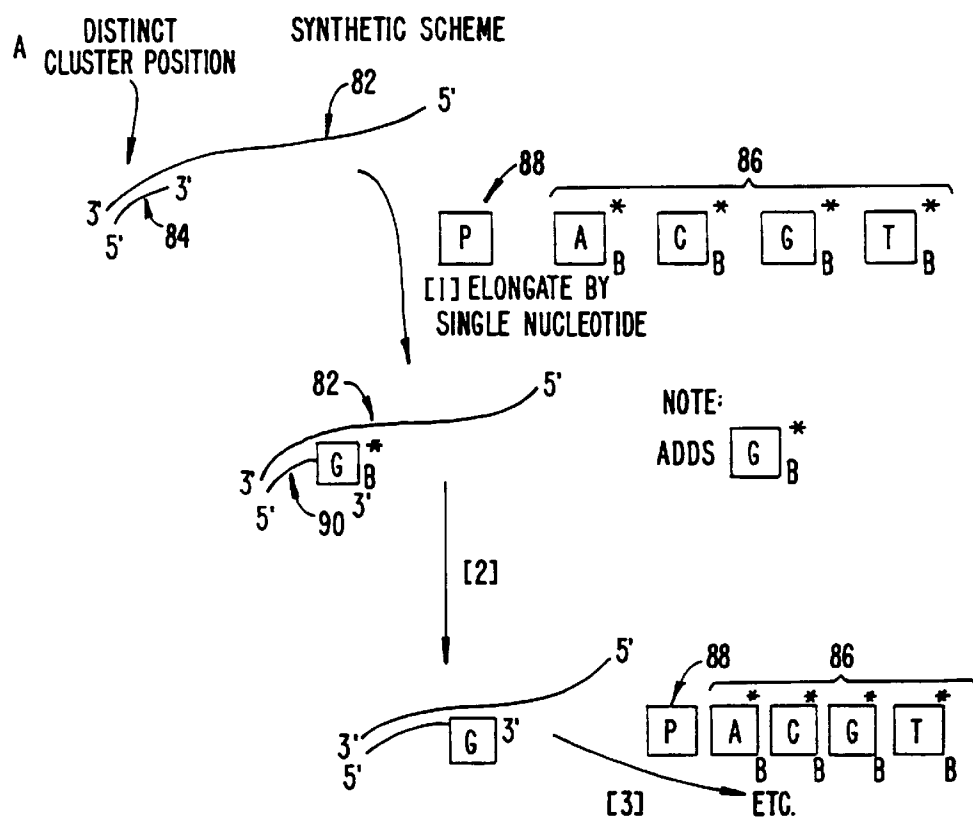
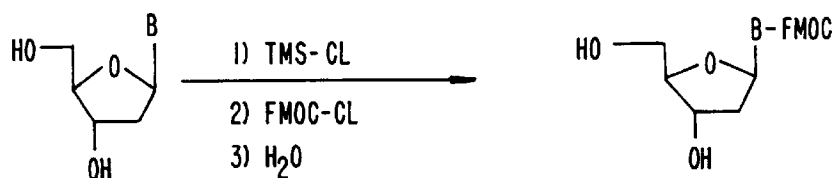


FIG. 8.

PATHWAY TO PROTECTED NUCLEOTIDES



PREFERRED PATHWAY TO BASE PROTECTION AND FUNCTIONALIZATION

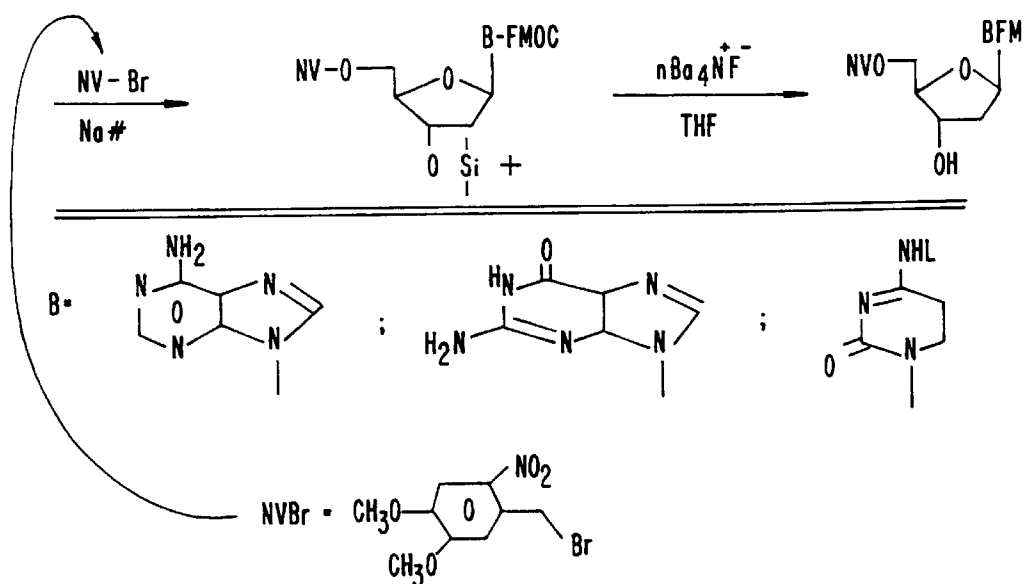
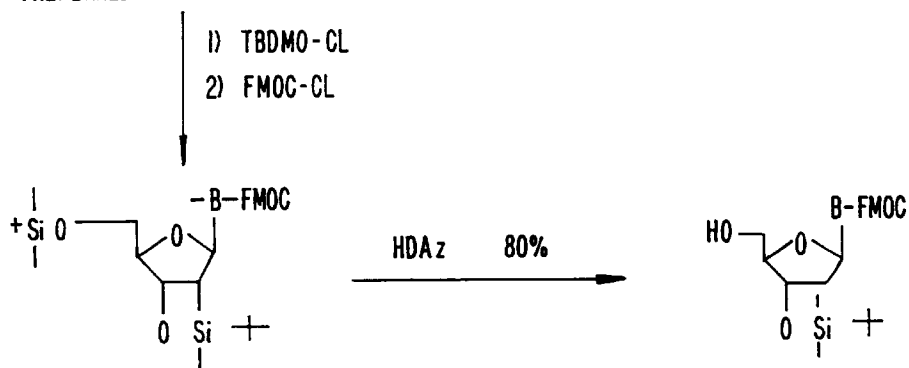


FIG. 9.

U.S. Patent

May 11, 1999

Sheet 9 of 10

5,902,723

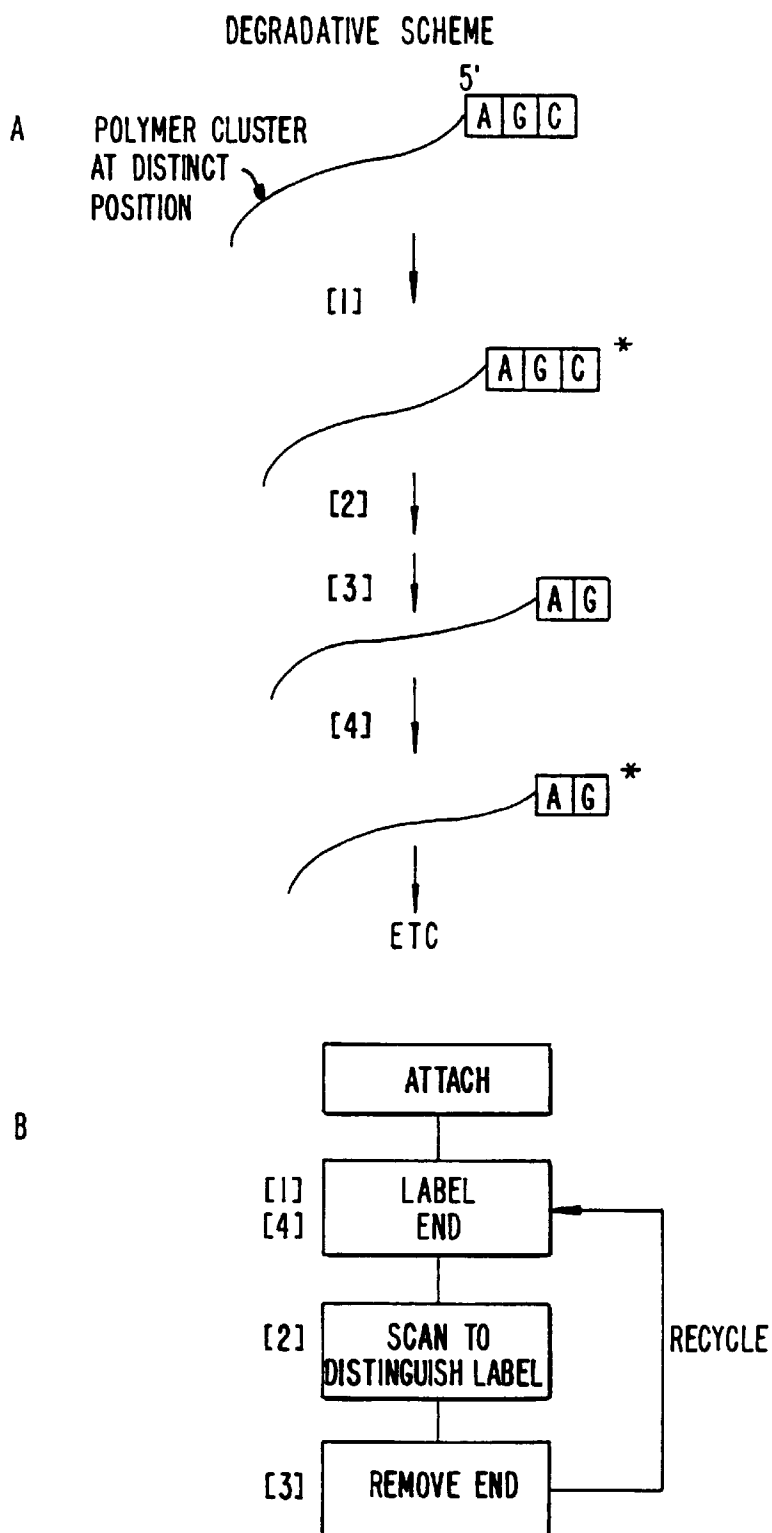


FIG. 10.

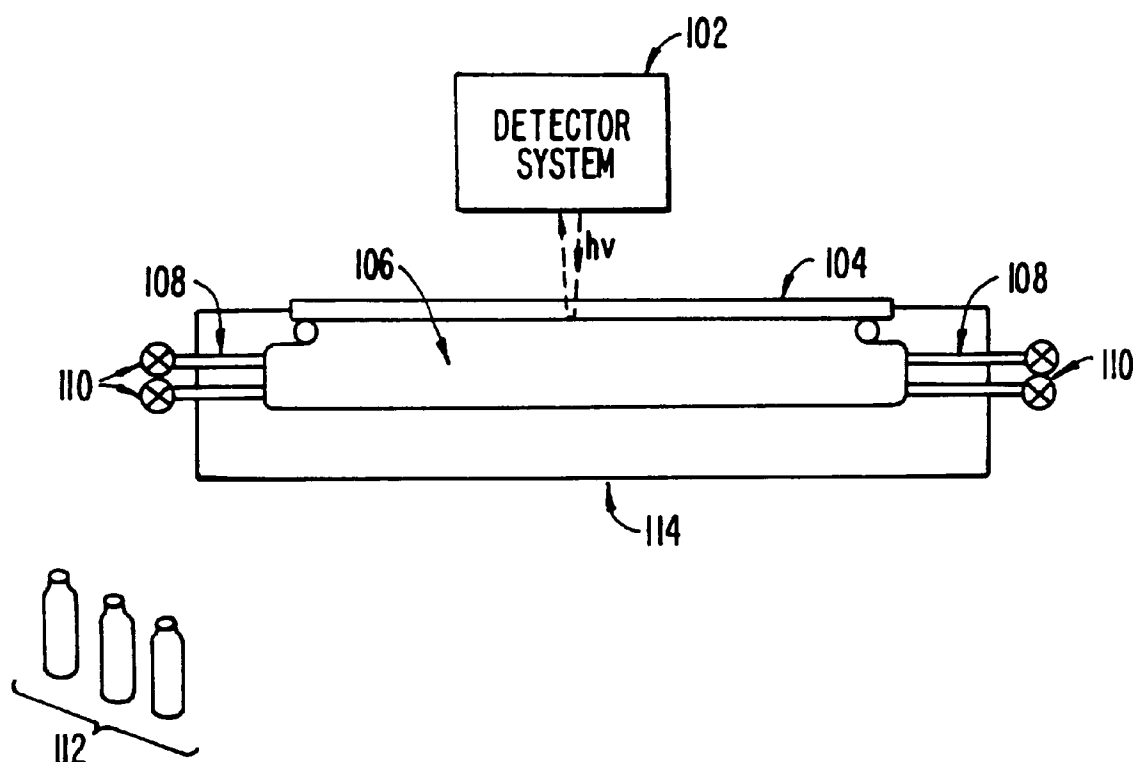


FIG. 11.

5,902,723

1

ANALYSIS OF SURFACE IMMOBILIZED POLYMERS UTILIZING MICROFLUORESCENCE DETECTION

CROSS REFERENCE TO RELATED APPLICATIONS

This application is a continuation under 37 CFR §1.60 of U.S. Ser. No. 07/626,730, filed Dec. 6, 1990, now U.S. Pat. No. 5,547,839, which is a continuation-in-part of U.S. Ser. No. 07/492,462, now U.S. Pat. No. 5,143,854, filed Mar. 7, 1990, which is a continuation-in-part of U.S. Ser. No. 07/362,901, filed Jun. 7, 1989, now abandoned. Related applications include U.S. Ser. No. 07/612,671, filed Nov. 1, 1990, now U.S. Pat. No. 5,252,743, which is a continuation-in-part of U.S. Ser. No. 07/435,316, filed Nov. 13, 1989, now abandoned; U.S. Ser. No. 07/624,120, filed Dec. 6, 1990, now abandoned; and U.S. Ser. No. 07/624,114, filed Dec. 6, 1990, now abandoned. Each of the above is incorporated herein by reference.

BACKGROUND OF THE INVENTION

The present invention relates to the determination of the sequences of polymers immobilized to a substrate. In particular, one embodiment of the invention provides a method and apparatus for sequencing many nucleic acid sequences immobilized at distinct locations on a matrix surface. The principles and apparatus of the present invention may be used, for example, also in the determination of sequences of peptides, polypeptides, oligonucleotides, nucleic acids, oligosaccharides, phospholipids and other biological polymers. It is especially useful for determining the sequences of nucleic acids and proteins.

The structure and function of biological molecules are closely interrelated. The structure of a biological polymer, typically a macromolecule, is generally determined by its monomer sequence. For this reason, biochemists historically have been interested in the sequence characterization of biological macromolecule polymers. With the advent of molecular biology, the relationship between a protein sequence and its corresponding encoding gene sequence is well understood. Thus, characterization of the sequence of a nucleic acid encoding a protein has become very important.

Partly for this reason, the development of technologies providing the capability for sequencing enormous amounts of DNA has received great interest. Technologies for this capability are necessary for, for example, the successful completion of the human genome sequencing project. Structural characterization of biopolymers is very important for further progress in many areas of molecular and cell biology.

While sequencing of macromolecules has become extremely important, many aspects of these technologies have not advanced significantly over the past decade. For example, in the protein sequencing technologies being applied today the Edman degradation methods are still being used. See, e.g., Knight (1989) "Microsequencers for Proteins and Oligosaccharides," *Bio/Technol.* 7:1075-1076. Although advanced instrumentation for protein sequencing has been developed, see, e.g., Frank et al. (1989) "Automation of DNA Sequencing Reactions and Related Techniques: A Work Station for Micromanipulation of Liquids," *Bio/Technol.* 6:1211-1213, this technology utilizes a homogeneous and isolated protein sample for determination of removed residues from that homogeneous sample.

Likewise, in nucleic acid sequencing technology, three major methods for sequencing have been developed, of which two are commonly used today. See, e.g., Sambrook et

2

al. (1989) *Molecular Cloning: A Laboratory Manual* (2d Ed.) Vols. 1-3, Cold Spring Harbor Press, New York, which is hereby incorporated herein by reference. The first method was developed by Maxam and Gilbert. See, e.g., Maxam and Gilbert (1980) "Sequencing End-Labeled DNA with Base-Specific Chemical Cleavages," *Methods in Enzymol.* 65:499-560, which is hereby incorporated herein by reference. The polymer is chemically cleaved with a series of base-specific cleavage reagents thereby generating a series of fragments of various lengths. The various fragments, each resulting from a cleavage at a specific base, are run in parallel on a slab gel which resolves nucleic acids which differ in length by single nucleotides. A specific label allows detection of cleavages at all nucleotides relative to the position of the label.

This separation requires high resolution electrophoresis or some other system for separating nucleic acids of very similar size. Thus, the target nucleic acid to be sequenced must usually be initially purified to near homogeneity.

Sanger and Coulson devised two alternative methods for nucleic acid sequencing. The first method, known as the plus and minus method, is described in Sanger and Coulson (1975) *J. Mol. Biol.* 94:441-448, and has been replaced by the second method. Subsequently, Sanger and Coulson developed another improved sequencing method known as the dideoxy chain termination method. See, e.g., Sanger et al. (1977) "DNA Sequencing with Chain-Termination Inhibitors," *Proc. Natl. Acad. Sci. USA* 74:5463-5467, which is hereby incorporated herein by reference. This method is based on the inability of 2', 3' dideoxy nucleotides to be elongated by a polymerase because of the absence of a 3' hydroxyl group on the sugar ring, thus resulting in chain termination. Each of the separate chain terminating nucleotides are incorporated by a DNA polymerase, and the resulting terminated fragment is known to end with the corresponding dideoxy nucleotide. However, both of the Sanger and Coulson sequencing techniques usually require isolation and purification of the nucleic acid to be sequenced and separation of nucleic acid molecules differing in length by single nucleotides.

Both the polypeptide sequencing technology and the oligonucleotide sequencing technologies described above suffer from the requirement to isolate and work with distinct homogeneous molecules in each determination.

In the polypeptide technology, the terminal amino acid is sequentially removed and analyzed. However, the analysis is dependent upon only one single amino acid being removed, thus requiring the polypeptide to be homogeneous.

In the case of nucleic acid sequencing, the present techniques typically utilize very high resolution polyacrylamide gel electrophoresis. This high resolution separation uses both highly toxic acrylamide for the separation of the resulting molecules and usually very high voltages in running the electrophoresis. Both the purification and isolation techniques are highly tedious, time consuming and expensive processes.

Thus, a need exists for the capability of simultaneously sequencing many biological polymers without individual isolation and purification. Moreover, dispensing with the need to individually perform the high resolution separation of related molecules leads to greater safety, speed, and reliability. The present invention solves these and many other problems.

SUMMARY OF THE INVENTION

The present invention provides the means to sequence hundreds, thousands or even millions of biological macro-

5,902,723

3

molecules simultaneously and without individually isolating each macromolecule to be sequenced. It also dispenses with the requirement, in the case of nucleic acids, of separating the products of the sequencing reactions on dangerous polyacrylamide gels. Adaptable to automation, the cost and effort required in sequence analysis will be dramatically reduced.

This invention is most applicable, but not limited, to linear macromolecules. It also provides specific reagents for sequencing both oligonucleotides and polypeptides. It provides an apparatus for automating the processes described herein.

The present invention provides methods for determining the positions of polymers which terminate with a given monomer, where said polymers are attached to a surface having a plurality of positionally distinct polymers attached thereto, said method comprising the steps of:

labeling a terminal monomer in a monomer type specific manner; and

scanning said surface, thereby determining the positions of said label. In one embodiment, the polymers are polynucleotides, and usually the labeling of the terminal marker comprises incorporation of a labeled terminal monomer selected from the group of nucleotides consisting of adenine, cytidine, guanidine and thymidine.

An alternative embodiment provides methods for concurrently determining which subset of a plurality of positionally distinct polymers attached to a solid substrate at separable locations terminates with a given terminal subunit, said method comprising the steps of:

mixing said solid substrate with a solution comprising a reagent, which selectively marks positionally distinct polymers which terminate with said given terminal subunit; and

determining with a detector which separable locations are marked, thereby determining which subset of said positionally distinct polymers terminated with said given terminal subunit. In one version, the solution comprises a reagent which marks the positionally distinct polymer with a fluorescent label moiety. In another version the terminal subunit is selected from the group consisting of adenosine, cytosine, guanosine, and thymine.

Methods are also provided for determining which subset of a plurality of primer polynucleotides have a predetermined oligonucleotide, wherein the polynucleotides are complementary to distinctly positioned template strands which are attached to a solid substrate, said method comprising the steps of:

selectively marking said subset of primer polynucleotides having the predetermined oligonucleotide; and

detecting which polynucleotides are marked. In one embodiment, the oligonucleotide subunit is a single nucleotide; in another the marking comprises elongating said primer with a labeled nucleotide which is complementary to a template; and in a further embodiment the marking step uses a polymerase and a blocked and labeled adenine.

The invention embraces methods for concurrently obtaining sequence information on a plurality of polynucleotides by use of a single label detector, said method comprising the steps of:

attaching a plurality of positionally distinct polynucleotides to a solid substrate at separable locations;

labeling said plurality of polynucleotides with a terminal nucleotide specific reagent, said label being detectable using said label detector;

4

determining whether said specific labeling reagent has labeled each separable location. Often, the labeling is performed with reagents which can distinguishably label alternative possible nucleotide monomers. One embodiment uses four replica substrates each of which is labeled with a specific labeling reagent for adenine, cytosine, guanine, or thymine. Usually, the labeling and determining steps are performed in succession using reagents specific for each of adenine, cytosine, guanine, and thymine monomers.

An alternative embodiment provides methods for concurrently obtaining sequence information on a plurality of polynucleotides, said method comprising the steps of:

attaching distinct polynucleotides to a plurality of distinct solid substrates;

labeling said plurality of solid substrates with a terminal nucleotide specific labeling reagent; and

determining whether said specific labeling reagent has labeled each distinct substrate. The method can be performed using a continuous flow of distinct solid substrates through a reaction solution.

A method is provided for simultaneously sequencing a plurality of polymers made up of monomer units, said plurality of polymers attached to a substrate at definable positions, said method comprising the steps of:

mixing said substrate with a reagent which specifically recognizes a terminal monomer, thereby providing identification among various terminal monomer units; and

scanning said substrate to distinguish signals at definable positions on said substrate; and

correlating said signals at defined positions on said substrate to provide sequential series of sequence determinations. Often, the plurality of polymers are synthesized by a plurality of separate cell colonies, and the polymers may be attached to said substrate by a carbonyl linkage. In one embodiment, the polymers are polynucleotides, and often the substrate comprises silicon. The scanning will often identify a fluorescent label. In one embodiment, the reagent exhibits specificity of removal of terminal monomers, in another, the reagent exhibits specificity of labeling of terminal monomers.

The invention also embraces methods for sequencing a plurality of distinctly positioned polynucleotides attached to a solid substrate comprising the steps of:

hybridizing complementary primers to said plurality of polynucleotides;

elongating a complementary primer hybridized to a polynucleotide by adding a single nucleotide; and

identifying which of said complementary primers have incorporated said nucleotide. In some versions, the elongating step is performed simultaneously on said plurality of polynucleotides linked to said substrate. Typically, the substrate is a two dimensional surface and the identifying results from a positional determination of the complementary primers incorporating the single defined nucleotide. A silicon substrate is useful in this method.

Methods, are provided where the linking is by photocrosslinking polynucleotide to said complementary primer, where said primer is attached to said substrate. The elongating will be often catalyzed by a DNA dependent polymerase. In various embodiments, a nucleotide will have a removable blocking moiety to prevent further elongation, e.g., NVOC.

5,902,723

5

A nucleotide with both a blocking moiety and labeling moiety will be often used.

A further understanding of the nature and advantages of the invention herein may be realized by reference to the remaining portions of the specification and the attached drawings.

BRIEF DESCRIPTION OF THE FIGURES

FIGS. 1A–D illustrates a simplified and schematized embodiment of a degradative scheme for polymer sequencing.

FIGS. 2A–D illustrates a simplified and schematized embodiment of a synthetic scheme for polymer sequencing.

FIG. 3 illustrates a coordinate mapping system of a petri plate containing colonies. Each position of a colony can be assigned a distinct coordinate position.

FIGS. 4A–C illustrates various modified embodiments of the substrates.

FIGS. 5A–B (Examiner's Amendment Feb. 21, 1996) illustrates an idealized scanning result corresponding to a particular colony position.

FIG. 6 illustrates particular linkers useful for attaching a nucleic acid to a silicon substrate. Note that thymine may be substituted by adenine, cytidine, guanine, or uracil.

FIG. 7 illustrates an embodiment of the scanning system and reaction chamber.

FIG. 8 illustrates the application of the synthetic scheme for sequencing as applied to a nucleic acid cluster localized to a discrete identified position. FIG. 8A illustrates schematically, at a molecular level, the sequence of events which occur during a particular sequencing cycle. FIG. 8B illustrates, in a logic flow chart, how the scheme is performed.

FIG. 9 illustrates the synthesis of a representative nucleotide analog useful in the synthetic scheme. Note that the Fmoc may be attached to adenine, cytosine, or guanine.

FIG. 10 illustrates the application of the degradative scheme for sequencing as applied to a nucleic acid cluster localized to a discrete identified position. FIG. 10A illustrates schematically, at a molecular level, the sequence of events which occur during a particular sequencing cycle. FIG. 10B illustrates in a logic flow chart how the scheme is performed.

FIG. 11 illustrates a functionalized apparatus for performing the scanning steps and sequencing reaction steps.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

I. Sequencing Procedure for a Generic Polymer

A. Overview

1. Substrate and matrix
2. Scanning system
3. Synthetic/degradative cycles
4. Label
5. Utility
- B. Substrate/Matrix
1. Non-distortable
2. Attachment of polymer
- C. Scanning system
1. Mapping to distinct position
2. Detection system
3. Digital or analog signal
- D. Synthetic or degradative cycle

6

1. Synthetic cycles
 - a. synthetic scheme
 - b. blocking groups
2. Degradative cycles
3. Conceptual principles
- E. Label

1. Attachment
2. Mode of detection
- F. Utility

II. Specific Embodiments

- A. Synthetic method
- B. Chain degradation method

III. Apparatus

I. Sequencing Procedure for a Generic Polymer

The present invention provides methods and apparatus for the preparation and use of a substrate having a plurality of polymers with various sequences where each small defined contiguous area defines a small cluster of homogeneous polymer sequences. The invention is described herein primarily with regard to the sequencing of nucleic acids but may be readily adapted to the sequencing of other polymers, typically linear biological macromolecules. Such polymers include, for example, both linear and cyclical polymers or nucleic acids, polysaccharides, phospholipids, and peptides having various different amino acids, heteropolymers in which the polymers are mixed, polyurethanes, polyesters, polycarbonates, polyureas, polyamides, polyethyleneimines, polyarylene sulfides, polysiloxanes, polyimides, polyacetates or mixed polymers of various sorts. In a preferred embodiment, the present invention is described in the use of sequencing nucleic acids.

Various aspects of the patents and applications in the cross reference above are applicable to the substrates and matrix materials described herein, to the apparatus used for scanning the matrix arrays, to means for automating the scanning process, and to the linkage of polymers to a substrate.

A. Overview

The present invention is based, in part, on the ability to perform a step wise series of reactions which either extend or degrade a polymer by defined units.

FIG. 1 schematizes a simplified linear two monomer polymer made up of A type and B type subunits. A degradative scheme is illustrated. Panel A depicts a matrix with two different polymers located at positions 10 and 14, but with no polymer linked at position 12. A reaction is employed to label all of these polymers at the terminus opposite the attachment of the monomer. Panel B illustrates a label (designated by an asterisk) incorporated at position 16 on the terminal monomers. A scan step is performed to locate positions 10 and 14 where polymers have been linked, but no polymer is located at position 12. The entire matrix is exposed to a reagent which is specific for removing single terminal A monomers, which are also labeled. The reagent is selected to remove only a single monomer; it will not remove further A monomers. Removal of the labeled A monomer leaves a substrate as illustrated in panel C. A scan step is performed and compared with the previous scan, indicating that the polymer located at position 10 has lost its label, i.e., that polymer at 10 terminated with an A monomer. The entire matrix is then exposed to a second reagent which is specific for removing terminal B monomers which are also labeled. Note that only a single B on each monomer is removed and that successive B monomers are not affected. Removal of the labeled B monomer leaves a substrate as illustrated in panel D. Another scan step is performed, indicating that the polymer located at position 14 has lost its label, i.e., it terminated with a B monomer. The sequence of

5,902,723

7

treatments and scans is repeated to determine the successive monomers. It will be recognized that if the labeled A and B are distinguishable, i.e., the label on polymers at sites 10 and 14 may be distinguished, a single removal step can be performed to convert the substrate as illustrated in panel B directly to that illustrated in panel D.

An alternative embodiment employs synthetic reactions where a synthetic product is made at the direction of the attached polymer. The method is useful in the synthesis of a complementary nucleic acid strand by elongation of a primer as directed by the attached polymer.

FIG. 2 illustrates a similar simplified polymer scheme, where the A and B monomer provide a complementary correspondence to A' and B' respectively. Thus, an A monomer directs synthetic addition of an A' monomer and a B monomer directs synthetic addition of a B' monomer. Panel A depicts monomers attached at locations 18 and 22, but not at location 20. Each polymer already has one corresponding complementary monomer A' or B. The matrix, with polymers, is subjected to an elongation reaction which incorporates, e.g., single labeled A' monomers 24 but not B' monomers, as depicted in panel B. The label is indicated by the asterisk. Note that only one A monomer is added. A scan step is performed to determine whether polymers located at positions 18 or 22 have incorporated the labeled A' monomers. The polymer at position 18 has, while the polymer at position 22 has not. Another elongation reaction which incorporates labeled B' monomers 26 is performed resulting in a matrix as depicted in panel C. Again note that only one, and not successive B' monomers, is added. Another scan is performed to determine whether a polymer located at sites 18 or 22 has incorporated a labeled B' monomer, and the result indicates that the polymer located at site 22 has incorporated the labeled B' monomer. A next step removes all of the labels to provide a substrate as depicted in panel D. As before, if the polymer which incorporated a labeled A' monomer is distinguishable from a polymer which incorporated a labeled B' monomer, the separate elongation reactions may be combined producing a panel C type matrix directly from a panel A type matrix and the scan procedure can distinguish which terminal monomer was incorporated.

It will be appreciated that the process may be applied to more complicated polymers having more different types of monomers. Also, the number of scan steps can be minimized if the various possible labeled monomers can be differentiated by the detector system.

Typically, the units will be single monomers, though under certain circumstances the units may comprise dimers, trimers, or longer segments of defined length. In fact, under certain circumstances, the method may be operable in removing or adding different sized units so long as the units are distinguishable. However, it is very important that the reagents used do not remove or add successive monomers. This is achieved in the degradative method by use of highly specific reagents. In the synthetic mode, this is often achieved with removable blocking groups which prevent further elongation.

One important aspect of the invention is the concept of using a substrate having homogeneous clusters of polymers attached at distinct matrix positions. The term "cluster" refers to a localized group of substantially homogeneous polymers which are positionally defined as corresponding to a single sequence. For example, a coordinate system will allow the reproducible identification and correlation of data corresponding to distinct homogeneous clusters of polymers locally attached to a matrix surface. FIG. 3 illustrates a mapping system providing such a correspondence, where

8

transfer of polymers produced by a colony of organisms to a matrix preserves spatial information thereby allowing positional identification. The positional identification allows correlation of data from successive scan steps.

In one embodiment, bacterial colonies producing polymers are spatially separated on the media surface of a petri plate as depicted in panel A. Alternatively, phage plaques on a bacterial lawn can exhibit a similar distribution. A portion of panel A is enlarged and shown in panel B. Individual colonies are labeled C1-C7. The position of each colony can be mapped to positions on a coordinate system, as depicted in panel C. The positions of each colony can then be defined, as in a table shown in panel D, which allows reproducible correlation of scan cycle results.

Although the preferred embodiments are described with respect to a flat matrix, the invention may also be applied using the means for correlating detection results from multiple samples after passage through batch or continuous flow reactions. For example, spatially separated polymers may be held in separate wells on a microtiter plate. The polymers will be attached to a substrate to retain the polymers as the sequencing reagents are applied and removed.

The entire substrate surface, with homogeneous clusters of polymers attached at defined positions, may be subjected to batch reactions so the entire surface is exposed to a uniform and defined sequence of reactions. As a result, each cluster of target polymers for sequencing will be subjected to similar reactive chemistry. By monitoring the results of these reactions on each cluster localized to a defined coordinate position, the sequence of the polymer which is attached at that site will be determined.

FIG. 4, panel A illustrates solid phase attached polymers linked to particles 32 which are individually sequestered in separate wells 34 on a microtiter plate. The scanning system will separately scan each well. FIG. 4 panel B illustrates marbles 36 to which polymers are attached. The marbles are automatically fed in a continuous stream through the reaction reagents 38 and past a detector 40. The marbles may be carefully held in tubes or troughs which prevent the order of the beads from being disturbed. In a combination of the two embodiments, each polymer is attached to a plurality of small marbles, and marbles having each polymer are separated, but retained in a known order. Each marble is, in batch with a number of analogous marbles having other polymers linked individually to them, passed through a series of reagents in the sequencing system. For example, A2, B2, and C2 are subjected to sequencing reactions in batch, with label incorporated only for the second monomer. A3, B3, and C3 are likewise treated to determine the third monomer. Likewise for A_n, B_n, and C_n. However, within each batch, the detection will usually occur in the order A, B, and C, thereby providing for correlation of successive detection steps for the A polymer beads, for the B polymer beads, and for the C polymer beads.

FIG. 5 illustrates a signal which might result from a particular defined position. Panel A illustrates the position of a given colony relative to the positions corresponding to the positional map. The scan system will typically determine the amount of signal, or type of signal, at each position of the matrix. The scan system will adjust the relationship of the detector and the substrate to scan the matrix in a controllable fashion. An optical system with mirrors or other elements may allow the relative positions of the substrate and detection to be fixed. The scanner can be programmed to scan the entire substrate surface in a reproducible manner, or to scan only those positions where polymer clusters have been localized. A digital data map, panel B, can be generated from the scan step.

5,902,723

9

Thus, instead of subjecting each individual and separated polymer to the series of reactions as a homogeneous sample, a whole matrix array of different polymers targeted for sequencing may be exposed to a series of chemical manipulations in a batch format. A large array of hundreds, thousands, or even millions of spatially separated homogeneous regions may be simultaneously treated by defined sequencing chemistry.

The use of a coordinate system which can reproducibly assay a defined position after each reaction cycle can be advantageously applied according to this invention. For example, a colony plaque lift of polymers can be transferred onto a nitrocellulose filter or other substrate. A scanning detector system will be able to reproducibly monitor the results of chemical reactions performed on the target polymers located at the defined locations of particular clones. An accurate positioning can be further ensured by incorporating various alignment marks on the substrate.

The use of a high resolution system for monitoring the results of successive sequencing steps provides the possibility for correlating the scan results of each successive sequencing reaction at each defined position.

The invention is dependent, in part, upon the stepwise synthesis or degradation of the localized polymers as schematized in FIGS. 1 and 2. The synthetic scheme is particularly useful on nucleic acids which can be synthesized from a complementary strand. Otherwise, a stepwise degradation scheme may be the preferred method. Although single monomer cycles of synthesis or degradation will usually be applicable, in certain cases the technology will be workable using larger segments, e.g., dimers or trimers, in the cyclic reactions.

The present invention also provides methods for production or selection of monomer-specific degradative reagents based upon catalytic antibody constructs. Antibody binding sites exhibiting specificity for binding particular terminal monomers can be linked to cleavage reagents or active sites of cleavage enzymes. Thus, reagents which are specific for particular terminal nucleotides may function to remove them in a specific fashion.

The invention also makes use of a means for detecting or labeling the polymers. Particular sequencing chemistry can be selected for specificity in reacting with terminal monomer units. Alternatively, indirect labeling methods may be applied which can distinguish between different terminal monomers. Another alternative scheme allows for terminal labeling which is not monomer-specific, but with the determination of the monomer based upon specificity of post-label reagents or upon monomer-distinguishable labels. Suitable such reagents will be antibodies or other reagents having specificity for distinguishing between different labeled terminal monomer residues and cleaving only those labeled monomer residues.

Thus, although neither the reaction nor the label need necessarily be specific, at least one of the pair must be specific. A comparison of label signal before and after a reaction allows determination of the change in label signal after monomer specific reactions are performed, and thereby provides the means to deduce the identity of the monomer at a given position.

B. Substrate/Matrix

The substrate or matrix has relatively few constraints on its composition. Preferably, the matrix will be inert to the sequencing reactions to which the polymers attached thereto will be subjected. Typically, a silicon or glass substrate will be used, but other suitable matrix materials include ceramics, or plastics, e.g., polycarbonate, polystyrene,

10

delrin, and cellulose, and any other matrix which satisfies these functional constraints.

In one embodiment, the matrix should be sufficiently nondeformable that the scanning system can reproducibly scan the matrix and reliably correlate defined positions with earlier and later scan operations. However, by including alignment markings on the substrate, the need for absolute rigidity of the substrate may be reduced.

In an alternative embodiment, the matrix may merely be large enough that the attached polymer may be separated from a liquid phase containing the sequencing reagents. In this embodiment, a single detection unit is used to analyze the label in a multiplicity of different samples after each of the reaction steps. Thus, different samples may be separably treated in distinct wells of a microtiter dish.

Separate homogeneous polymers can be introduced to solid phase beads in each microtiter well. Sequencing reagents may be individually introduced separately into each well, or transferred from well to well with the polymers remaining in the correct well due to their solid phase attachments.

In an alternative approach, the solid phase matrix may be marbles or other particularly shaped articles. Spherical shapes, solid or hollow, are preferred because they can be easily transported through troughs or tubing which retains their relative orders. By feeding a succession of beads through appropriate reaction baths and past a detector in a known and retained order, a succession of label detection results from a bead may be correlated and converted into a polymer sequence.

The attachment of the homogeneous clusters of target polymers to the substrate can be achieved by appropriate linkage chemistry. As indicated before, the linkage should be stable and insensitive to the sequencing reagents used. The specific linkages will depend, of course, upon the particular combination of substrate and polymer being used.

Typically, the most useful chemical moieties which will be used are amines. Typical substrate derivatized groups include aminopropyl triethoxysilane, hydroxypropylacrylate, or hydroxy reagents, see, e.g., Ser. No. 624,120, filed Dec. 6, 1990, now abandoned. Typical polymer derivatized groups include nitroveratryl and nitroveratryl oxycarbonyl. Linkage types are also illustrated and detailed in Ser. No. 624,120 and Ser. No. 624,114, each filed Dec. 6, 1990, both now abandoned.

FIG. 6 illustrates one preferred linkage chemistry for nucleic acids. An NVO-derivatized Ser. No. 624,120, filed Dec. 6, 1990, now abandoned. The specific conditions for synthesis of thymidine are described therein and are adaptable to other nucleotides and nucleosides. The nucleoside analog is further derivatized with an appropriate R group at the 3' hydroxyl. Preferred R groups are indicated in FIG. 6. The linkage produces a photosensitive blocked nucleoside suitable for phosphoramidite synthesis of further polynucleotides which can serve as a complementary strand for hybridization of other polymers. The hybrids of the complementary strands may be covalently crosslinked using acridine dyes or other intercalative reagents, e.g., psoralen. See, e.g., Kornberg (1980) *DNA Replication* Freeman, San Francisco; Wiesehahn, et al. (1978) *Proc. Natl. Acad. Sci. USA* 75:2703-2707, and Sheldon (1986) U.S. Pat. No. 4,582,789 which are each incorporated herein by reference.

The linkage should be substantially inert to the cyclic sequencing reactions and scan cycles. Usually, the linkage will be at a defined and homogeneous polymer position, preferably at the end opposite where the sequencing chemistry takes place. Although the type of linkage is dependent

5,902,723

11

upon the polymer being sequenced, various types of polymers have preferred linkages. For polypeptides, amino terminal or carboxyl terminal linkages will be preferred. Specific amino terminal linkages include amino butyric acid, amino caproic acids, and similar carboxylic acids. Specific carboxyl terminal linkages include butyric acid, caproic acid, and other carboxylic acids, hydrocarbon, and ethers. See now abandoned Ser. No. 435,316, filed Nov. 13, 1989, and allowed Ser. No. 492,462, filed Mar. 7, 1990, now U.S. Pat. No. 5,143,854 which are incorporated herein by reference. For nucleic acids, the linkages will typically be either 5' or 3' linkages. Suitable 3' linkages include those illustrated in FIG. 6, and others described in copending Ser. No. 624,114, filed Dec. 6, 1990, now abandoned.

Alternatively, for complementary polymers, particularly nucleic acids, linkage may be via crosslinkage of the complementary polymers where the complementary strand is directly attached to the matrix. Acridine dyes, e.g., psoralen, or a similar crosslinking agent between the strands can be used. See, e.g., Dattagupta, et al., "Coupling of Nucleic Acids to Solid Support By Photochemical Methods," U.S. Ser. No. 4,713,326; and U.S. Ser. No. 4,542,102; and Chatterjee, M. et al. (1990) *J. Am. Chem. Soc.* 112:6397; which describe useful crosslinking reagents, and are hereby incorporated herein by reference.

For polynucleotides, the preferred attachment to the matrix is through a synthetic oligomer by the 5' end of each target sequence. This oligomer is designed to anneal to the desired target templates used in a synthetic system or to the polynucleotide used in the degradation approach. In one embodiment, a vector sequence which is complementary to the immobilized oligonucleotide is incorporated adjacent the cloning inserts, thereby providing a common complementary sequence for each insert. In particular, a cloning vector will be selected with a defined sequence adjacent the insert. See, e.g., Sambrook et al., (1989) *Molecular Cloning: A Laboratory Manual*, Vols. 1-3, Cold Spring Harbor Press, which is hereby incorporated herein by reference. This defined sequence is used, in some embodiments, as a common linker for all of the vector inserts. The inserts, adjacent to this linker, will be transferable by hybridization to the matrix linked complementary sequences. The hybrids are crosslinked by addition of a suitable crosslinker under appropriate conditions, for example, photocrosslinking by psoralen with uv light. See, e.g., Song et al. (1979) *Photochem. Photobiol.* 29:1177-1197; Cimino et al. (1985) *Ann. Rev. Biochem.* 54:1151-1193; and Parsons (1980) *Photochem. Photobiol.* 32:813-821; each of which is incorporated herein by reference. Using these approaches, the oligonucleotide linker serves as both the attachment linker and the polymerization primer.

FIG. 6 illustrates a preferred 3' terminal linkage designed for a phosphoramidite linkage of a synthetic primer and the reactions forming them. The chemical reactions for actually performing the linkage will be similar to those used for oligonucleotide synthesis instruments using phosphoramidite or similar chemistry. Applied Biosystems, Foster City, Calif. supplies oligonucleotide synthesizers.

C. Scanning System

The scanning system should be able to reproducibly scan the substrate. Where appropriate, e.g., for a two dimensional substrate where the polymers are localized to positions thereon, the scanning system should positionally define the clusters attached thereon to a reproducible coordinate system. It is important that the positional identification of clusters be repeatable in successive scan steps. Functionally, the system should be able to define physical positions to a

12

coordinate system as described above and illustrated in FIGS. 3 and 4.

In alternative embodiments, the system can operate on a cruder level by separately detecting separate wells on a microtiter plate, or by scanning marbles which pass by the detector in an embodiment as described above and illustrated in FIG. 4.

The scanning system would be similar to those used in electrooptical scanning devices. See, e.g., the allowed Ser. No. 492,462, filed Mar. 7, 1990, now U.S. Pat. No. 5,143,854 and copending Ser. No. 624,120, filed Dec. 6, 1990, now abandoned. The system could exhibit many of the features of photographic scanners, digitizers or even compact disk reading devices. For example, a model no. PM500-A1 x-y translation table manufactured by Newport Corporation can be attached to a detector unit. The x-y translation table is connected to and controlled by an appropriately programmed digital computer such as an IBM PC/AT or AT compatible computer. The detection system can be a model no. R943-02 photomultiplier tube manufactured by Hamamatsu, attached to a preamplifier, e.g., a model no. SR440 manufactured by Stanford Research Systems, and to a photon counter, e.g., an SR430 manufactured by Stanford Research System, or a multichannel detection device. Although a digital signal may usually be preferred, there may be circumstances where analog signals would be advantageous.

The stability and reproducibility of the positional localization in scanning will determine, to a large extent, the resolution for separating closely positioned polymer clusters in a 2 dimensional substrate embodiment. Since the successive monitoring at a given position depends upon the ability to map the results of a reaction cycle to its effect on a positionally mapped cluster of polymers, high resolution scanning is preferred. As the resolution increases, the upper limit to the number of possible polymers which may be sequenced on a single matrix will also increase. Crude scanning systems may resolve only on the order of 100 μm , refined scanning systems may resolve on the order of 100 μm , more refined systems may resolve on the order of about 10 μm with optical magnification systems a resolution on the order of 1.0 μm is available, and more preferably a resolution on the order of better than 0.01 μm is desired. The limitations on the resolution may be diffraction limited and advantages may arise from using shorter wavelength radiation for the photo-optical deprotection fluorescent scanning steps. However, with increased resolution, the time required to fully scan a matrix will be increased and a compromise between speed and resolution will necessarily be selected. Parallel detection devices which will provide high resolution with shorter scan times will be applicable where multiple detectors will be moved in parallel.

With other embodiments, resolution often is not so important and sensitivity might be emphasized. However, the reliability of a signal may be pre-selected by counting photons and continuing to count for a longer period at positions where intensity of signal is lower. Although this will decrease scan speed, it can increase reliability of the signal determination. Various signal detection and processing algorithms may be incorporated into the detection system, such as described in copending Ser. No. 624,120, filed Dec. 6, 1990, now abandoned. In one embodiment, the distribution of signal intensities of pixels across the region of signal are evaluated to determine whether the distribution of intensities corresponds to a time positive signal.

The detection system for the signal or label will depend upon the label used, which may be defined by the chemistry

5,902,723

13

available. For optical signals, a combination of an optical fiber or charged couple device (CCD) may be used in the detection step. In those circumstances where the matrix is itself transparent to the radiation used, it is possible to have an incident light beam pass through the substrate with the detector located opposite the substrate from the polymers. For electromagnetic labels, various forms of spectroscopy systems can be used. Various physical orientations for the detection system are available and discussion of important design parameters is provided, e.g., in Jovin, *Adv. in Biochem. Biophys.*, which is hereby incorporated herein by reference.

Various labels which are easily detected include radioactive labels, heavy metals, optically detectable labels, spectroscopic labels and the like. Various photoluminescent labels include those described in Ser. No. 624,114, filed Dec. 6, 1990, now abandoned. Protection and deprotection are described, e.g., in McCray, et al. (1989) *Ann. Rev. Biophysical Chemistry* 18:239-270, and in Ser. No. 624,120, filed Dec. 6, 1990, now abandoned, each of which is hereby incorporated herein by reference.

With a processing system, the speed of scanning may be dramatically increased with a system which only scans positions where known clusters of polymer are attached. This allows the scanning mechanism to skip over areas which have been determined to lack any polymer clusters and avoids loss of time in scanning useless regions of the matrix. Moreover, various problems with spurious or overlapping signals may be adjusted for by appropriate analysis.

A scanning apparatus which may be used for the presently described uses is schematically illustrated in FIG. 7. A substrate **52** is placed on an x-y translation table **54**. In a preferred embodiment the x-y translation table is a model no. PM500-A1 manufactured by Newport Corporation. The x-y translation table is connected to and controlled by an appropriately programmed digital computer **56** which may be, for example, an appropriately programmed IBM PC/AT or AT compatible computer. Of course, other computer systems, special purpose hardware, or the like could readily be substituted for the AT computer used herein for illustration. Computer software for the translation and data collection functions described herein can be provided based on commercially available software including, for example, "Lab Windows" licensed by National Instruments, which is incorporated herein by reference for all purposes.

The substrate and x-y translation table are placed under a microscope **58** which includes one or more objectives **60**. Light (about 488 nm) from a laser **62**, which in some embodiments is a model no. 2020-05 argon ion laser manufactured by Spectraphysics, is directed at the substrate by a dichroic mirror **64** which passes greater than about 520 nm wavelength light but reflects 488 nm light. Dichroic mirror **64** may be, for example, a model no. FT510 manufactured by Carl Zeiss. Light reflected from the mirror then enters the microscope **58** which may be, for example, a model no. Axioscop 20 manufactured by Carl Zeiss. Fluorescently marked materials on the substrate will fluoresce >488 nm light, and the fluoresced light will be collected by the microscope and passed through the mirror. The fluorescent light from the substrate is then directed through a wavelength filter **66** and, thereafter through an aperture plate **68**. Wavelength filter **66** may be, for example, a model no. OG530 manufactured by Melles Griot and aperture plate **68** may be, for example, a model no. 477352/477380 manufactured by Carl Zeiss.

The fluoresced light then enters a photomultiplier tube **70** which in one embodiment is a model no. R943-02 manu-

14

factured by Hamamatsu, the signal is amplified in preamplifier **72** and photons are counted by photon counter **74**. The number of photons is recorded as a function of the location in the computer **56**. Pre-Amp **72** may be, for example, a model no. SR440 manufactured by Stanford Research Systems and photon counter **74** may be a model no. SR430 manufactured by Stanford Research Systems. The substrate is then moved to a subsequent location and the process is repeated. In preferred embodiments the data are acquired every 1 to 100 μm with a data collection diameter of about 0.8 to 10 μm preferred. In embodiments with sufficiently high fluorescence, a CCD detector with broadfield illumination is utilized.

By counting the number of photons generated in a given area in response to the laser, it is possible to determine where fluorescently marked molecules are located on the substrate. Consequently, for a substrate which has a matrix of polypeptides, for example, synthesized on the surface thereof, it is possible to determine which of the polypeptides has incorporated a fluorescently marked monomer.

According to preferred embodiments, the intensity and duration of the light applied to the substrate is controlled by varying the laser power and scan stage rate for improved signal-to-noise ratio by maximizing fluorescence emission and minimizing background noise. Signal analysis may improve the resolution and reliability of the system. The time of photon counting may be varied at various positions to provide high signal to background or noise.

D. Synthetic or Degradative Cycle

The present invention provides a substrate with positionally separated polymers for sequencing. The separation may be by solid phase carriers separated in separate wells, by separately manipulable carriers such as beads or marbles, or by physical separation of regions on a two-dimensional substrate surface. Each cluster region is a target for the sequencing reactions. Although the reactions are, in various embodiments, performed on all the clusters together, each cluster can be individually analyzed by following the results from the sequence of reactions on polymer clusters at positionally defined locations.

The synthetic mode, as illustrated in FIG. 1 is easily applied to the sequencing of nucleic acids, since one target strand may serve as the template to synthesize the complementary strand. The nucleic acid can be DNA, RNA or mixed polymers. For the purposes of illustration, and not by limitation, the sequencing steps for DNA are described in detail. The synthetic mode, an example of which is depicted in FIG. 8 for nucleotides, may also be useful in circumstances where synthesis occurs in response to a known polymer sequence. The synthetic scheme depends, in part, on the stepwise elongation by small and identifiable units. A polymerase is used to extend a primer complementary to a target template. The primer is elongated one nucleotide at a time by use of a particular modified nucleotide analog to which a blocking agent is added and which prevents further elongation. This blocking agent is analogous to the dideoxy nucleotides used in the Sanger and Coulson sequencing procedure, but in certain embodiments here, the blockage is reversible. This analog is also labeled with a removable moiety, e.g., a fluorescent label, so that the scanning system can detect the particular nucleotide incorporated after its addition to the polymerization primer.

Panel 4A illustrates the cycle of sequence reactions in one embodiment. The template polymer **82** located at a particular site has already been linked to substrate. The template **82** and complementary primer **84** are hybridized. Often, the primer **84** is common to all of the target template sequences,

5,902,723

15

selected by its common occurrence on a selected cloning vector. The primer **84** is also often covalently crosslinked to the target template **82** using psoralen and U.V. light.

Labeled and blocked monomers **86** are shown, the label depicted by the asterisk and the polymerization blocking groups indicated by B. A compatible polymerase **88** which can elongate the primer with the labeled blocked monomers **86** is used in reaction **1**. In the preferred embodiment, the separate labeled monomers can be distinguished from one another by the wavelength of fluorescent emission.

In the example illustrated, a labeled blocked guanosine monomer has been incorporated into the elongated primer **90**.

Step 2 is a scan, where the signal at the position corresponding to template **82** indicates that the guanosine analog was incorporated. Reaction **2** is performed, which removes both the label and blocking group. It will be recognized that the blocking group prevents elongation by any more than a single nucleotide in each reaction cycle. Reaction **3** is equivalent to reaction **1**, though the substrate primer has been elongated by one monomer.

Panel B illustrates the scheme in a logic flow chart. The template **82** is attached to the substrate, either directly or through the primer. Reaction **1** elongates the primer by a single labeled blocked nucleotide. A scan step is performed and the blocking and labeling agents are removed. The elongation reaction is performed and the cycle repeated.

For a nucleic acid, a unit for addition would typically be a single nucleotide. Under certain circumstances, dimers or trimers or larger segments may be utilized, but a larger number of different possible nucleotide elements requires high distinguishability in other steps. For example, there are only four different nucleotide monomer possibilities, but there are sixteen different dimer possibilities. The distinction among four possibilities is more precise and simple than among sixteen dimer possibilities. To prevent elongation by a unit length greater than one monomer, the nucleotide should be blocked at the position of 3' elongation. Usually, the nucleotide will be blocked at the 3' hydroxyl group where successive nucleotides would be attached. In contrast to a dideoxy nucleotide, typically the blocking agent will be a reversible blocking agent thereby allowing for deblocking and subsequent elongation.

Variations may be easily incorporated into the procedure. If the labels on the monomers are not distinguishable, successive substrate scans can be performed after each monomer is provided with conditions allowing its incorporation. Alternatively, a small fraction of permanently blocked but reversibly labeled monomers may be incorporated. Those specific molecules which incorporate the blocked monomers are permanently removed from further polymerization, but such is acceptable if the labeling moiety is also removed.

1. other monomers

One important functional property of the monomers is that the label be removable. The removal reaction will preferably be achieved using mild conditions. Blocking groups sensitive to mild acidic conditions, mild basic conditions, or light are preferred. The label position may be anywhere on the molecule compatible with appropriate polymerization, i.e., complementary to the template, by the selected polymerase. A single polymerase for all of the modified nucleotide is preferred, but a different polymerase for each of the different monomers can be used.

Nucleotide analogs used as chain-terminating reagents will typically have both a labeling moiety and a blocking agent while remaining compatible with the elongation enzy-

16

mology. As the blocking agent will usually be on the 3' hydroxyl position of the sugar on a nucleotide, it would be most convenient to incorporate the label and the blocking agent at the same site, providing for a single reaction for simultaneous removal of the label and blocking agent. However, it is also possible to put a label on another portion of the nucleotide analog than the 3' hydroxyl position of the sugar, thereby requiring a two-step reaction cycle for removing the blocking and labeling groups.

Analogous will be found by selecting for suitable combinations of appropriate nucleotides with compatible polymerases. In particular, it is desired that a selected polymerase be capable of incorporating a nucleotide, with selectivity, having both the blocking moiety and the label moiety attached. It has been observed that RNA polymerases are less fastidious with respect to the nucleotide analogues which will be polymerized into a growing chain. See, e.g., Rozovaskaya, T., et al. (1977) *Molekulyarnaya Biologiya*, 11:598-610; Kutateladze, T., et al. (1986) *Molekulyarnaya Biologiya*, 20:267-276; and Chidgevadze, Z., et al. (1985) *FEBS Letters*, 183:275-278. Moreover, those references also indicate that rather significant chemical moieties may be attached at the 2' or 3' positions on a nucleotide, and still be correctly incorporated at the growing chain terminus.

In particular, it is not necessary that the same nucleotide have both the reversible blocking moiety and the removable labeling moiety, as a combination of two separate nucleotide analogues could be utilized, e.g., N1, which is reversibly blocked and not labeled, and N2, which is irreversibly blocked but removably labeled. Note that the removal of label may be affected by destruction of the label, e.g., fluorescence destruction, or preferably by removal. Both of these nucleotides might be, for instance, A analogues. With the mixture, at an appropriate sequence position of a target sequence, N1 and N2 nucleotides can be incorporated at an appropriate ratio, and these can be polymerized by either two separate polymerases, or preferably a single polymerase.

For example, two separate polymerases might be necessary, P1 which incorporates N1, and P2 which incorporates N2. At the given location in the sequence, some of the growing polymers will incorporate N1 with P1 polymerase, and others will incorporate N2 with the P2 polymerase. The proportions of N1, N2, P1, and P2 may be titrated to get the desired fractional proportions of the N1 reversibly blocked nucleotides and the N2 labeled but irreversibly blocked nucleotides.

As all of the growing chains have blocked nucleotides, no elongation takes place beyond a single nucleotide. The N2 nucleotides provide a specific label, detected in the scanning step. After determination of the incorporated label, the label may be removed or destroyed, and those irreversibly terminated growing chains become permanently removed from further participation in the sequencing process. Photodestruction may be achieved by a high intensity laser beam of the correct wavelength. See, e.g., March (1977) *Advanced Organic Chemistry: Reactions, Mechanisms and Structure* (2d Ed) McGraw; and Carey and Sundberg (1980) *Advanced Organic Chemistry: part A Structure and Mechanisms*, Plenum.

Next, the reversible blocking moiety is removed, providing a new set of slightly longer polymers ready for the next step. Of course, the amount of label necessary to be incorporated must be detectable, preferably with a clear, unambiguous positive signal. The amount of label incorporated will depend, in part, upon the conditions in the polymerizing step and the relative incorporation of the N1 and N2 nucle-

5,902,723

17

otides. The proportions of the nucleotides, polymerases, and other reagents may be adjusted to appropriately incorporate the desired proportions of the nucleotides.

In an embodiment where a single polymerase will incorporate both N1 and N2, the relative proportions and conditions to get the correct incorporation levels of the two nucleotides can be titrated. In an alternative preferred embodiment, a single nucleotide will have both the removable label and the reversible blocking moiety.

A similar approach may be necessary where only some fraction of the nucleotide analogues is labeled. Separate polymerases might also be useful for such situations, and each polymerase may have special conditions necessary for activity.

Procedures for selecting suitable nucleotide and polymerase combinations will be readily adapted from Ruth et al. (1981) *Molecular Pharmacology* 20:415-422; Kutateladze, T., et al. (1984) *Nuc. Acids Res.*, 12:1671-1686; Kutateladze, T., et al. (1986) *Molekulyarnaya Biologiya* 20:267-276; Chidgeavadze, Z., et al. (1985) *FEBS Letters*, 183:275-278; and Rozovskaya, T., et al. (1977) *Molekulyarnaya Biologiya* 11:598-610.

The determination of termination activity is done in two steps. First, nucleotide analogues are screened for the ability of the compound to inhibit polymerase activity. Then the nucleotide analogue is tested for base-specific termination as manifested by generating a correct DNA sequencing ladder on a template of known sequence. The appropriate reaction conditions are those used for conventional sequencing reactions with the respective polymerases. The conditions are then modified in the usual ways to obtain the optimal conditions for the particular terminator compound (e.g. concentration of terminator, ratio of terminator to dNTP, Mg⁺⁺, and other reagents critical to proper polymerase function.

By way of example, an approach employing the polymerase known as reverse transcriptase (AMV) will be described. The initial conditions are essentially as described by Prober, et al. (1987) *Science* 238:336-341.

A nucleotide analogue is first selected from the group available from a commercial source such as Amersham, New England Nuclear, or Sigma Chemical Company. In particular, nucleotides which are reversibly blocked from further elongation, especially at the 5' or 3' —OH will be used.

General properties which are desired have been described. Each of these analogs can be tested for compatibility with a particular polymerase by testing whether such polymerase is capable of incorporating the labeled analog. Various polymerases may be screened, either natural forms of the mentioned types, or variants thereof. Polymerases useful in connection with the invention include *E. Coli* DNA polymerase (Klenow fragment); and Klenow and Henningsen (1970) *Proc. Nat'l Acad. Sci. USA* 65:168-175; and Jacobsen et al. (1974) *Eur. J. Biochem.* 45:623-627; modified and cloned versions of T7 DNA polymerase (SequenaseTM and Sequenase 2.0TM); see Tabor and Richardson (1987) *Proc. Nat'l Acad. Sci. USA* 84:4767-4771; and Tabor and Richardson (1987) *J. Biol. Chem.* 262:15330-15333; Taq DNA polymerase from thermostable *Thermus aquaticus*; see Chien et al. (1976) *J. Bacteriol.* 127:1550-1557; and its cloned version AmplitaqTM; Saiki and Gelfand (1989) *Amplifications* 1:4-6; T4 DNA polymerase; see Nossal (1974) *J. Biol. Chem.* 249:5668-5676, and various reverse transcriptases, both RNA- and DNA- dependent DNA polymerases, e.g., avian retroviruses; see Houts (1979) *J. Virology* 29:517-522; and murine retroviruses; see

18

Kotewicz et al. (1985) *Gene* 35:249-258; Gerard et al. (1986) *DNA* 5:271-279; and Bst polymerase; see Ye, S. and Hong (1987) *Scientia Sinica* 30:503-506.

In order to ensure that only a single nucleotide is added at a time, a blocking agent is usually incorporated onto the 3' hydroxyl group of the nucleotide. Optimally, the blocking agent should be removable under mild conditions (e.g., photosensitive, weak acid labile, or weak base labile groups), thereby allowing for further elongation of the primer strand with a next synthetic cycle. If the blocking agent also contains the fluorescent label, the dual blocking and labeling functions will be achieved without the need for separate reactions for the separate moieties.

The blocking group should have the functional properties of blocking further elongation of the polymer. Additional desired properties are reversibility and inertness to the sequencing reactions. Preferably, where an enzymatic elongation step is used, the monomers should be compatible with the selected polymerase. Specific examples for blocking groups for the nucleic acids include acid or base labile groups at the 3'OH position. See, e.g., Gait (1984) *Oligonucleotide Synthesis: A Practical Approach*, IRL Press, Oxford.

A DNA-dependent DNA polymerase is the polymerase of choice. Polymerases used for conventional DNA sequencing, for example, Klenow fragment of *E. coli* DNA Pol, Sequenase (modified T7 DNA polymerase), Taq (*Thermus aquaticus*) DNA polymerase, Bst (*Bacillus stearothermophilus*), DNA polymerase, reverse transcriptase (from AMV, MMLV, RSV, etc.) or other DNA polymerases will be the polymerases of choice. However, there is a functional constraint that the polymerase be compatible with the monomer analogues selected. Screening will be performed to determine appropriate polymerase and monomer analog combinations.

Removal of the blocking groups may also be unnecessary if the labels are removable. In this approach, the chains incorporating the blocked monomers are permanently terminated and will no longer participate in the elongation processes. So long as these blocked monomers are also removed from the labeling process, a small percentage of permanent loss in each cycle can also be tolerated.

The fluorescent label may be selected from any of a number of different moieties. The preferred moiety will be a fluorescent group for which detection is quite sensitive. Various different fluorescence-labeling techniques are described, for example, in Kambara et al. (1988) "Optimization of Parameters in a DNA Sequenator Using Fluorescence Detection," *Bio/Technol.* 6:816-821; Smith et al. (1985) *Nucl. Acids Res.* 13:2399-2412; and Smith et al. (1986) *Nature* 321:674-679, each of which is hereby incorporated herein by reference. Fluorescent labels exhibiting particularly high coefficients of destruction may also be useful in destroying nonspecific background signals.

Appropriate blocking agents include, among others, light sensitive groups such as 6-nitoveratryloxycarbonyl (NVOC), 2-nitobenzoyloxycarbonyl (NBOC), α,α -dimethyldimethoxybenzyloxycarbonyl (DDZ), 5-bromo-7-nitroindolyl, o-hydroxy-2-methyl cinnamoyl, 2-oxymethylene anthraquinone, and t-butyl oxycarbonyl (TBOC). Other blocking reagents are discussed, e.g., in U.S. Ser. No. 07/492,462; Patchornik (1970) *J. Amer. Chem. Soc.* 92:6333; and Amit et al. (1974) *J. Org. Chem.* 39:192, all of which are hereby incorporated herein by reference. Additional blocking agents attached to particular positions may be selected according to the functional directives provided herein.

5,902,723

19

FIG. 9 schematically illustrates the synthesis of a generic protected nucleotide. A suitable nucleotide is labeled with the FMOC fluorescently detectable label by reaction under the conditions described, e.g., in copending Ser. No. 624, 114, filed Dec. 6, 1990, now abandoned, with (TMS-Cl), FMOC-Cl, and H₂O. A protection moiety will be added using conditions also described there.

Various nucleotides possessing features useful in the described method can be readily synthesized. Labeling moieties are attached at appropriate sites on the nucleotide using chemistry and conditions as described, e.g., in Gait (1984) *Oligonucleotide Synthesis*. Blocking groups will also be added using conditions as described, e.g., in copending Ser. No. 624,114, filed Dec. 6, 1990, now abandoned. FIG. 9 also outlines various reactions which lead to useful nucleotides.

Additionally, the selected polymerases used in elongation reactions should be compatible with nucleotide analogs intended for polymerization to the primer. Simple screening procedures for nucleotide and polymerase combinations may be devised to verify that a particular combination is functional. A test using primer with template which directs the addition of the nucleotide analog to be incorporated will determine whether the combination is workable. Natural polymerases or variants thereof may be used under particular defined conditions.

The degradative scheme is generally illustrated in FIG. 1, an example more generally applicable to biological macromolecular polymers is depicted in FIG. 10. This method is useful for a wider variety of polymers without the limitations imposed by the need to replicate the polymer. The degradative sequencing technique depends, in part, upon the ability to specifically label or distinguish between various different terminal monomers at particular matrix positions. Reactions for specific removal of a defined monomer unit are important.

This monomer distinguishability can arise from an ability to differentiate between label on the various possible monomers in the polymer. As a second means, distinguishability can come from specific reagents which react with particularity on different monomers. Thus, for instance, labels may be used which generally attach to the terminal nucleotide, but whose fluorescent signal differs depending upon the nucleotide. As a third means, a reagent which specifically affects the label on only one monomer may be used, as described below.

In the first example, every polymer cluster will be labeled at a particular end, e.g., the 5' end, without specificity for the monomer located there. The scan step will be able to distinguish the terminal monomers, after which each labeled terminal monomer is specifically removed. The general label step is repeated in the cycle as described.

In the second means for distinguishability, reagents are used which produce a signal which is dependent upon the terminal nucleotide. For example, a labeling molecule which binds only to one specific terminal monomer will provide a monomer specific label. This will provide a cycle much like the first means for distinguishability where the properties of the label is different depending upon the terminal nucleotide to which each specific labeling reagent binds.

In the third means for distinguishability, an individual reagent labels or affects only a specific terminal monomer. Polymers susceptible to each reagent by virtue of terminating with the corresponding monomer will have their label specifically affected. A scan of the matrix after each step and comparison with the earlier scans will determine which positions correspond to polymers ending with a susceptible monomer. Performing a removal step with a second

20

monomer-specific reagent followed by a scan will identify those positional locations having polymer clusters ending with that second monomer. A similar reagent for the other possible monomers will further define all of the possibilities. Finally, when all of the possible monomers have been removed, the labeling reaction may be repeated and the succession of specific reagent and scanning steps will also be repeated. This procedure allows for a succession of automated steps to determine the sequence of the polymer clusters localized to distinct positions.

Finally, a combination of both specificity of reagent and ability to distinguish label on different monomers can be utilized. Neither alone need be relied upon exclusively. Thus, in the case of nucleotides, an ability to distinguish into two separate classes of nucleotides, e.g., A and C from G and T, combined with specific reagents for distinguishing between the indistinguishable label pairs, e.g., in the example provided, A from C, or G from T, can also provide sufficient information for sequencing.

Instead of performing four specific reactions on the same substrate matrix, each of the four individual reactions can be performed on separate parallel matrices. Four separate substrate matrices may be made by a replica plating or successive transfers, each matrix having the same spatial distribution of polymer clusters. Thus, each separate substrate can be subjected to only a single specific reagent in a highly optimized reaction. On each cycle, one out of the four parallel substrates should show a signal indicating the monomer at the terminal for the cluster at a given matrix position.

Likewise, two parallel substrates can be provided, and each of the parallel substrates is used to determine two of the four possible nucleotides at each position. Instead of treating a single matrix with four separate reactions, this approach allows treating each of two substrates with only two separate reactions. By minimizing the number of reactions to which each chip is exposed, the side reactions will be minimized, the chemistry will be optimized, and the number of cycles through which a matrix will survive will be optimized. This provides an advantage in the number of cycles to which a matrix can be subjected before the signal to noise becomes indistinguishable.

E. Label

The label is important in providing a detectable signal. The signal may be distinguishable among the various monomers by the nature of the signal, e.g., wavelength or other characteristic, as described in Prober et al. (1987) *Science* 238:336-341. A monomer-specific reagent can allow determination of whether each position has a particular terminal monomer by the presence or loss of label.

The label on the monomer may be attached by a noncovalent attachment, but will be preferably attached by a direct covalent attachment. The label will typically be one which is capable of high positional resolution and does not interfere with the nucleotide-specific chemistry or enzymology. Although many different labels may be devised including enzyme linked immunosorbent assays (ELISA), spectrophotometric labels, light producing or other labels, a fluorescent moiety is the preferred form. For example, an avidin/biotin type affinity binding may be useful for attaching a particular label. Alternatively, an antibody may be used which is specific for binding to a particular terminal monomer. A wide variety of other specific reagents can be used to provide a labeling function. See, for example, copending Ser. No. 624,114, filed Dec. 6, 1990, now abandoned, which is hereby incorporated herein by reference.

The means of detection utilized will be selected in combination with various other considerations. In some

5,902,723

21

circumstances, a spectroscopic label may be most compatible with a particular monomer. Enzyme linked assays with a spectrophotometric detection system are a workable system. Phosphorescent or light producing assays provide high sensitivity using charged couple devices. Fluorescent systems provide the same advantages, especially where the incident light beam is a laser. The fluorescent label also may provide the added advantage of fluorescing at different wavelengths for the different monomers, providing a convenient means to distinguish between different monomers. Other forms of label may be desired for various reasons, for example, magnetic labels, radioactive labels, heavy metal atoms, optically detectable labels, spectroscopically detectable labels, fluorescent labels, and magnetic labels.

For sequencing nucleic acids by this method, the labeled monomers are simpler than those monomers used for the synthetic method. The blocking group is unnecessary, but terminal specific reagents are more difficult to produce.

The preferred attachment sites will be at the same location as the blocking site, so a combined label and blocking moiety is more preferred. The label will be attached as described, e.g., in copending Ser. No. 624,114, filed Dec. 6, 1990, now abandoned.

Two types of degradation cycles can be used, either non-specific removal of the terminal labeled nucleotide, or a base-specific removal. With the non-specific removal means, each of the end monomers, when labeled, should be distinguishable from the other three monomer possibilities. This allows for determination of the terminal nucleotide for the cluster localized at a given matrix position. Then the terminal, labeled nucleotides are non-specifically removed and the newly exposed terminal nucleotides will be again distinguishably labeled.

By this scheme, a specific label for each of the different nucleotides may be provided. For example, fluorescent reagents specific for each of the nucleotides may provide a signal with a different wavelength. This will more usually occur when the fluorescent probe is located near the base moiety of the nucleotide. In the scanning step, the regions terminating with each of the four different nucleotides may be determined. Then, a reaction is performed that removes the labeled terminal nucleotides from all of the polymers. This removal may be either enzymatic, using a phosphatase, an exonuclease or other similar enzyme, or chemical, using acid, base, or some other, preferably mild, reagent. Again, the reactions are performed which label each of the terminal nucleotides and a scan step repeated in the same manner.

In the base-specific removal scheme, nucleotide-specific removal can be performed. For example, an enzyme which will function to remove only a single modified nucleotide, e.g., a 5'-fluorescein-dAMP-specific exonuclease, is constructed. This may be achieved by proper construction of a catalytic antibody. Other similar reagents may be generated for each of the other labeled nucleotide monomers.

Catalytic or derivatized antibodies to catalyze the removal of the 3'-end or 5'-most fluorescent base in a base-specific manner may be constructed as follows. A recombinant antibody library or a series of monoclonal antibodies is screened with fluorescent donor-quencher substrates. These substrates consist of a fluorescent labeled base (A, C, G, or T) on the 5' or 3' end joined by a 5' to 3' phosphodiester linkage to a second base. A collection of all four possible second bases for each of the four end bases gives the best selection target for the required non-specificity with respect to the second base. The second base is then tethered to an acceptor group in sufficient proximity to quench the fluorescence of the end group. In the presence of a catalytic

22

antibody with cleaving activity, a fluorescent signal occurs from the separation of the quenching group from the terminal fluorescent label. To assure both base and end specificity, the positive monoclonal antibody clones are rescreened against the other substrates.

Upon selection of an antibody exhibiting the desired specificity (or lack thereof), the reactive group for cleavage may be attached. This cleavage reagent may be chemical or enzymatic and will be attached by an appropriate length linker to the antibody binding site in an orientation which is consistent with the steric requirements of both binding and specific cleavage.

Particularly useful specific reagents may be produced by making antibodies specific for each of the four different modified terminal nucleotide bases. These antibodies would then specifically bind only to polymers terminating in the appropriate base analog. By combining a cleavage reagent to the specific antibody, a terminal nucleotide specific cleavage reagent is generated.

In one example of the degradative embodiment, all of the polymers may be uniformly labeled at a particular end. Thereafter, a specific removal reaction which removes only a particular nucleotide may be performed, leaving the three other nucleotides labeled. Thereafter, a scanning step is performed through which all regions which had incorporated that particular nucleotide will have lost the label through specific removal. Then, the second specific reagent will be applied which specifically removes the second labeled nucleotide, and the scanning step following that reaction will allow determination of all regions which lose the second particular nucleotide. This process is repeated with reagents specific for each of the last two remaining labeled nucleotides interspersed with scanning steps, thereby providing information on regions with each of the nucleotides located there. Then, the entire process may be repeated by labeling the next terminal nucleotides uniformly. As mentioned below, replication techniques may allow for making four separate but identical matrix substrates. Each substrate may be subjected to single nucleotide-specific reactions, and the scan results correlated with each of the other parallel substrates.

In the degradation scheme, the polynucleotide linkage to the matrix must be more carefully selected such that the free end of the oligonucleotide segments used for attachment will not interfere with the determinations of the target sequence terminus.

F. Utility

The present sequencing method is useful to monitor and check the accuracy and reliability of the synthetic processes described in now abandoned Ser. No. 362,901, filed Jun. 7, 1989, and U.S. Pat. No. 5,143,854. The present method can be used to check the final products synthesized therein, or to label each monomers as they are added stepwise to monitor the efficiency and accuracy of those synthetic methods.

The present invention can also be used to monitor or sequence matrix bound clusters of positionally distinct polymers. This sequencing process provides the capability of simultaneously sequencing a large plurality of distinct polymers which are positionally segregated.

The method will be used to sequence extremely large stretches of polymer, e.g., nucleic acids. A large number of shorter segments of a large sequence can be sequenced with alignment of overlaps either randomly generated, or in an ordered fashion, or particular sequenceable segments of a large segment can be generated. In one approach, a large segment is subcloned into smaller segments and a sufficient number of the randomly generated subclones are sequenced

5,902,723

23

as described herein to provide sequence overlap and ordering of fragments.

In an alternative approach, a large segment can be successively digested to generate a succession of smaller sized subclones with ends separated by defined numbers of monomers. The subclones can be size sorted by a standard separation procedure and the individual samples from a separation device manually or automatically linked to a matrix in a defined positional map. Fractions resulting from size separation can be spatially attached at defined positions, often at adjacent positions. Then polymer sequences at adjacent positions on the matrix will also be known to have ends which differ by, e.g., approximately 25 or 50 or more monomers, thereby providing significantly greater confidence in overlapping sequence data.

III. Specific Embodiments

A specific series of reactions for sequencing a matrix of polynucleotides is described.

A. Synthetic Method

This method involves annealing a primer (common to all the attached sequences by virtue of the cloning construction) near to the 3' end of the unknown target sequences. DNA polymerase, or a similar polymerase, is used to extend the chains by one base by incubation in the presence of dNTP analogs which function as both chain terminators and fluorescent labels. This is done in a one-step process where each of the four dNTP analogs is identified by a distinct dye, such as described in Prober et al. *Science* 238:336-341, or in four steps, each time adding one of the four bases, interspersed with a scanning identification step. When each cluster incorporates the proper one of the four bases and the fluorescence scanning is complete, the matrix is stripped of the label and the chain terminators are deblocked for a next round of base addition. Because the base addition is directed by the template strand, the complementary sequence of the fragments at each address of the matrix is deduced.

(1) Attachment to a surface.

Both degradative and synthetic sequencing methods begin by obtaining and immobilizing the target fragments of unknown sequence to be determined at specific locations on the surface.

There are several strategies for photo-directed attachment of the DNA strands to the surface in an orientation appropriate for sequencing. A caged biotin technique, see, e.g., Ser. No. 435,316, filed Nov. 13, 1989, and Ser. No. 612,671, filed Nov. 13, 1990, is available. Another technique that is especially applicable for the enzymatic synthesis method is to chemically attach a synthetic oligomer by the 3' end to the entire surface (see FIG. 6), to activate it for photocrosslinking (with psoralen, for example) and to anneal the complementary strands and photocrosslink the target strand of unknown sequence (complementary to this oligonucleotide at the 3' end) at the specific location addressed by light. In this case, the oligonucleotide serves as both the attachment linker and as the synthetic primer. A third method is to physically transfer individual nucleic acid samples to selected positions on the matrix, either manually or automatically.

Many sequences in each step are attached by cloning the library into a series of vectors identical except for the sequences flanking the insert. These primers can be added at the point of amplification of the cloned DNA with chimeric primers.

Alternatively, sequences are attached to a matrix substrate by colony or phage immobilization. This directly transfers the positional distribution on a petri plate to a usable substrate. Colonies representing a shotgun collection of

24

sequences (enough to assure nearly complete coverage by overlap) are spread over (or in) a nutrient surface at a density to give about 100 or more colonies or plaques in several square centimeters, and the colonies are allowed to grow to about 0.1 mm in diameter (the maximum possible density of clusters at this size is $\sim 10,000$ colonies/cm²). As described above, replica platings or successive transfers may allow for preparation of multiple matrices with identical positional distributions of polymers. Each separate matrix may then be dedicated to the reactions applicable to a single monomer.

For example, in the use of a phage library, on a petri dish, the transfer substrate surface is treated to release DNA from the phage. This is done, e.g., with CHCl₃ vapor, SDS-NaOH, or by heating. Prior to release of DNA, the phage particles are often adsorbed to the surface by way of an antibody to the coat protein that has been immobilized on the surface. This strategy prevents diffusion of the phage from the colonies. The matrix surface is prepared by coating with an oligonucleotide, immobilized to the surface by one end that has homology with the phage vector DNA adjacent to the cloning site.

The matrix surface is juxtaposed to the growth surface, and the phage DNA is allowed to anneal to the immobilized oligonucleotide. The growth surface is removed, and the hybrid is stabilized by psoralen or an equivalent crosslinking reagent.

This method provides an efficient one-step method of placing many DNA fragments onto the detection surface in preparation for sequencing. Although the colonies are not placed in predefined locations, the random arrangement of the clusters allows the final sequence to be assembled from correlation of overlap sequence data derived from sequence data derived from each of the defined positions of each target cluster.

Sequences are, in other embodiments, attached by a manual or automated transfer technique. A few cells from each colony in a library are toothpicked into microliter wells. The plate is heated to $\sim 100^\circ$ C. for a short period to lyse the cells and release the DNA. The plate is cooled and reagents for cyclic amplification of the DNA using, e.g., PCR technology, are added, including primers common to all the cloned sequences. See, e.g., Innis et al. (1990) *PCR Protocols: A Guide to Methods and Applications*, Academic Press, which is hereby incorporated herein by reference. The DNA is amplified asymmetrically by unbalanced primer concentration to yield an excess of one strand for sequencing and attached to a substrate by manual or automated means.

An alternative form of automated localization is described above in positioning of a succession of smaller sized polymers which are manually or automatically linked to the substrate in a pattern reflecting sequence overlaps.

(2) Enzymatic polymerization method.

The nucleic acid template is, in some embodiments, attached to the surface by either the 5' or the 3' end, usually by a method as described above. A preferred method of attachment is to anneal the template to an oligonucleotide attached to the surface and to crosslink the template to the oligonucleotide. Oligonucleotide primers are usually synthesized chemically. In this case, the immobilized oligonucleotide may also serve as a primer for polymerization. Because polymerization proceeds 5' to 3' on the primer, the template will be attached by its 3' end, or a site 3' proximal to the region to be sequenced, for the purposes of the description to follow.

Step 1: A DNA-dependent, DNA polymerase such as those used for conventional DNA sequencing, for example, Klenow fragment of *E. coli* DNA Pol, Sequenase™

5,902,723

25

(modified T7 DNA polymerase), Taq (*Thermus aquaticus*) DNA polymerase, Bst (*Bacillus stearothermophilus*), DNA polymerase, reverse transcriptase (from AMV, MMLV, RSV, etc.) or other DNA polymerases, and the reaction components appropriate to the particular DNA polymerase selected, are placed in the incubation chamber in direct contact with the surface.

Step 2: Fluorescent chain terminators (analogues of dATP, dCTP, dGTP, and TTP, each labeled with fluorophore preferably emitting at a distinguishable wavelength) are added to the reaction at a sufficient concentration and under suitable reaction conditions (time, temperature, pH, ionic species, etc., see Sambrook et al. (1989) *Molecular Cloning*, vols. 1-3, and Prober et al.) to cause essentially all of the chains on the surface to be extended by one base and thereby terminated. Detection of the specific label thereby incorporated into each chain identifies the last base added at each positional address in the matrix.

Step 3: The chain termination should be reversible by some means, such as treatment with light, heat, pH, certain other chemical or biological (enzymatic) reagents, or some combination of these. Typically the chain termination results from a blocking moiety which is labile to mild treatment. By one of these means, the blocked 3'OH of the terminating base must be made available for chain extension in the next round of polymerization.

Step 4: There are several suitable labeled, terminator structures as follows:

- (a) The fluorophore itself functions as the chain terminator by placement on the 3' hydroxyl through a linkage that is easily and efficiently cleaved (removing the label and leaving the free 3'OH) by light, heat, pH shift, etc. The surface is scanned with a scanning system, e.g., the fluorescence detection system described in allowed Ser. No. 492,462, filed Mar. 7, 1990, now U.S. Pat. No. 5,143,854; and copending Ser. No. 624,120, filed Dec. 6, 1990, now abandoned. Then, preferably in a single step, the fluorophore is removed and the chain is activated for the next round of base addition.
- (b) The fluorophore is placed in a position other than the 3'OH of the nucleoside, and a different group is placed on the 3'OH of the dNTPs to function as a chain terminator. The fluorophore and the 3' blocking group are removed by the same treatment in a single step (preferably), or they may be removed in separate steps.
- (c) An alternative polymer stepwise synthetic strategy can be employed. In this embodiment, the fluorophores need not be removable and may be attached to irreversible chain terminators. Examples of such compounds for use in sequencing DNA include, but are not limited to, dideoxynucleotide triphosphate analogs as described by Prober et al. (1987) *Science* 238:336-341. A second, unlabeled and reversible, set of terminators is also required. Examples of these compounds are deoxynucleotide triphosphates with small blocking groups such as acetyl, tBOC, NBOC and NVOC on the 3'OH. These groups are easily and efficiently removed under conditions of high or low pH, exposure to light or heat, etc. After each round of base addition and detection, the fluorophores are deactivated by exposure to light under suitable conditions (these chains have their labeling moiety destroyed and remain terminated, taking part in no further reactions). The unlabeled, reversible terminators are unblocked at the 3'OH by the appropriate treatment to allow chain extension in subsequent rounds of elongation. The proportion of chains labeled in each round can be controlled by the concen-

26

tration ratio of fluorescent to non-fluorescent terminators, and the reaction can be driven to completion with high concentrations of the unlabeled terminators.

- (d) A single dye strategy is used where all the base analog terminators carry the same fluorophore and each is added one at a time: A, C, G, T. The addition of each base is followed by scanning detection. After all four fluorophores are added, reversal of the termination is performed, allowing for the addition of the next base analog. Then, each scanning step determines whether the immediately preceding labeled nucleotide had been incorporated at each distinct position.

The structures of the fluorescently labeled and reversible terminator base analogs are selected to be compatible with efficient incorporation into the growing chains by the particular DNA polymerase(s) chosen to catalyze extension. For example, where two different chain terminators are used, they may be utilized by two different polymerases that are both present during the chain extension step.

Step 5: An optional step is the permanent capping of chain extension failures with high concentrations of dideoxynucleotide triphosphates. This step serves to reduce the background of fluorescence caused by addition of an incorrect base because of inefficient chain extension (termination) at an earlier step.

Step 6: After scanning to determine fluorescence, the fluorophore is removed or deactivated. Deactivation of the fluorophore can be achieved by a photodestruction event. The chain elongation block is reversed (usually by removing a blocking group to expose the 3'OH) by suitable methods that depend on the particular base analogs chosen; and the substrate is washed in preparation for the next round of polymerization.

Step 7: Repeat the cycle.

B. Chain Degradation Method

This method involves labeling the last base of the chain (distal to the surface attachment) with a fluorescent tag followed by base-specific removal. All the polynucleotide clusters on the matrix are labeled using a standard labeling moiety. Base-specific removal of the last base of each chain, interspersed with fluorescence scanning of the array, will reveal the disappearance of fluorescence and hence the identity of the last base of each chain. When all four labeled end bases have been removed, the polymers attached to the matrix are relabeled and the process is repeated, working successively on the DNA chains.

Alternatively, if the label allows distinguishing between different monomers, simpler degradation processes may be employed. A single scan step can distinguish between all four possible terminal nucleotides. The four separate removal steps are then combined into a single nonspecific terminal nucleotide removal step.

The DNA will usually be attached to the substrate by the 3' or 5' terminus depending on the scheme of labeling and cleavage. Because there are well-known 5'-labeling methods, see, e.g., Gait (1984) *Oligonucleotide Synthesis: A Practical Approach*, IRL Press, Oxford, this discussion will assume the 3' end is attached to the substrate with the 5' end free.

Step 1: All the 5'-end bases are labeled with 5'-specific chemistry, e.g., 5' amino linkage to FITC, Nelson et al. (1989) *Nucl. Acids Res.* 17:7179-7186, which is hereby incorporated herein by reference.

Step 2: Scan the matrix to obtain the background level.

Step 3: Optional: Cap all of the labeling failures, e.g., polymers whose ends were not labeled.

5,902,723

27

Step 4: The terminal A's are removed with end-base, A-specific reagents (such a reagent may be chemical or biological). One example is a 5'-fluorescein-dAMP-specific exonuclease made as a catalytic antibody (see the description above for a scheme of producing this reagent).

Step 5: Scan the matrix to detect those chains that had terminated in A (these will be reduced in fluorescence compared to the fluorescent by labeled background).

Step 6: Repeat steps 4 and 5 for each of other three possible bases using the appropriate fluorescein-base-specific cleavage reagent and scan after removal of each of the C's, the G's, and the T's. This succession of steps will allow the determination of the terminal nucleotide of each positionally defined cluster.

Step 7: Relabel the 5' terminal nucleotide of all the new end bases that have been exposed by the earlier rounds of cleavage, and repeat the stepwise removal and scanning processes.

This approach can be extended to protein sequencing using 20 catalytic antibodies (or other amino acid-specific cleavage reagents), each recognizing a terminal amino acid and removing that terminal residue.

The process for sequencing may be summarized as follows for enzymatic polymerization:

- 1) Target DNA templates (to be sequenced) are attached at positionally defined locations on the matrix substrate.
- 2) Fluorescent chain terminators are added to a primer under conditions where all polymer chains are terminated after addition of the next base complementary to the template.
- 3) The matrix is scanned to determine which base was added to each location. This step correlates the added base with a position on the matrix.
- 4) Chains failing to extend (and therefore to terminate) are capped.
- 5) The fluorophores are removed or deactivated.
- 6) The terminators are activated for further chain extension, usually by removal of a blocking group.
- 7) Steps 2 through 6 are repeated to obtain the base-by-base sequence of many different positionally separated DNA fragments simultaneously.

C. Screening for new nucleotide analog/polymerase combinations.

The use of a functional combination of blocked nucleotide with a polymerase is important in the synthetic embodiment of the present invention. It is important to ensure that only a single nucleotide is incorporated at the appropriate step. The following protocol describes how to screen for a functional combination.

Test 1. (test for polymerase inhibition)

In a reaction volume of 20 μ l, mix

1 μ g M13mp19 single stranded DNA template

2.5 ng standard M13 primer (17-mer:5'-GTTTTCCCAGTCACGAC-3'

60 mM tris-Cl pH 8.5

7.5 mM MgCl₂

75 mM NaCl

Template and primer are annealed by heating to 95° C., then cooling to ~25° C.

Extension components are added:

50 AM (each) dATP, dCTP, dGTP, TTP

10 μ Ci P32 dATP

0.01 μ M to 1 mM of the putative terminator compound, further titrations may be desired.

20 units AMV reverse transcriptase water to 20 μ l final volume

28

The reaction is run at 42° C. for about 30 minutes.

Aliquots are taken at 10, 20, 30 minutes, and samples are TCA precipitated after the addition of 10 μ g tRNA carrier.

The filters are counted for acid-precipitable radioactivity and the mass of dATP incorporated is calculated as a function of reaction time.

Control reactions are run in parallel consisting of

A) no added terminator

B) 10 μ M and 100 μ M

The termination activity of the experimental samples relative to that of ddNTPs is estimated, and a nucleotide is appropriate for further testing if it substantially decreases the number of acid precipitable counts at any time or relative concentration.

Test 2 (test for base specific termination activity)

Reactions are run essentially as described by Prober et al. except:

1. Unlabelled primer is used

2. 1 μ Ci P32 dATP is included

3. No dideoxynTPs are added to the experimental samples (control reactions containing ddNTP at the usual concentrations, and no test terminators are run in parallel)

4. The test compound is added at a concentration estimated to give 1% and 10% inhibition of incorporation as determined by test #1.

The reactions are run for 10 min at 42° C. 100 μ M dNTPS are added and the reaction run for an additional 10 min. A portion of the reaction is prepared and run on a sequencing gel in the usual fashion. The ladders obtained with the test compound are compared with these obtained in the ddNTP reactions and the fidelity of the termination activity of the test compound is thereby assessed.

IV. Apparatus

The present invention provides a new use for an apparatus comprising a reaction chamber and a scanning apparatus which can scan a substrate material exposed to the chamber. FIG. 11 illustrates a system and a schematized reaction chamber to which is attached a silicon or glass substrate. The system has a detection system 102 as illustrated, in one embodiment, in FIG. 7. A silicon substrate 104, is attached against and forming a seal to make a reaction chamber 106. Leading into and out of the chamber are tubes 108, with valves 110 which control the entry and exit of reagents 112 which are involved in the stepwise reactions. The chamber is held at a constant temperature by a temperature block 114.

All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference. The invention now being fully described, it will be apparent to one of ordinary skill in the art that many changes and modifications can be made thereto without departing from the spirit or scope of the claims.

What is claimed is:

1. A method of analyzing a plurality of nucleic acids, comprising:

providing a plurality of different nucleic acid sequences immobilized upon a solid support, each of said different sequences being attached to said solid support in a distinct position;

treating said plurality of nucleic acid sequences with a catalytic antibody that is capable of catalyzing removal of a specific nucleotide in a nucleotide-specific manner to selectively remove a single terminal nucleotide;

identifying positions on said solid support in which a terminal nucleotide has been removed; and

5,902,723

29

repeating said treating and identifying steps to analyze said sequences of said plurality of different nucleic acids.

2. The method of claim 1, wherein said nucleic acids are labeled on said terminal nucleotide prior to said treating step, and said identifying step comprises identifying positions on said solid support in which a labeled terminal nucleotide has been removed.

3. The method of claim 2, wherein said label is a fluorescent label.

4. The method of claim 2, wherein each type of terminal nucleotide is labeled with a separate distinguishable label.

5. The method of claim 1, wherein said solid support is a planar substrate.

6. The method of claim 1, wherein said plurality of different nucleic acids are a plurality of different single stranded sequences.

7. A method of polynucleotide analysis, comprising:

(a) contacting at least one polynucleotide with a surface of a solid substrate, the surface having attached thereto in spatially defined locations a plurality of different oligonucleotide primers of defined nucleotide sequence and length, wherein the oligonucleotides are patterned on the surface of the substrate such that the oligonucleotide primers with different nucleotide sequences are located in different locations;

(b) hybridizing the oligonucleotide primers to at least one region of complementary nucleotide sequence contained in the at least one polynucleotide in a reaction mixture comprising a nucleic acid polymerase and a labelled nucleotide under conditions whereby the at least one hybridized primer is elongated by template-directed addition of the labelled nucleotide; and

(c) determining the location of the at least one hybridized primer from the location of label, the location of the hybridized primer indicating its sequence and thereby the complementary sequence of the at least one polynucleotide.

30

8. The method of claim 7, wherein the labeled nucleotide is a single type of nucleotide selected from the group consisting of adenine, cytidine, guanine, and thymidine.

9. The method of claim 7, wherein the labelled nucleotide is adenine, cytidine, guanine and thymidine, each of which bears a distinct label.

10. The method of claim 7, wherein a plurality of different nucleotides are contacted with said surface of said substrate in said contacting step, whereby a plurality of hybridized primers are elongated by template-directed addition of the labelled nucleotide; and the locations of the plurality of hybridized primers are determined from the locations of label, the locations of the hybridized primers indicating their sequence and thereby the complementary sequences of the plurality of different polynucleotides.

11. The method of claim 10, wherein the plurality of polynucleotides are obtained by digestion of a larger polynucleotide.

12. A method for determining the incorporation of nucleotides or polynucleotides comprising:

providing a nucleic acid array, said array having associated therewith at least two different nucleic acids, each of said different nucleic acids being associated with said array in a different defined position;

hybridizing at least a portion of a target nucleic acid to said nucleic acid array;

contacting said nucleic acid array with a plurality of nucleotides or polynucleotides and a template dependent nucleic acid polymerase;

determining the incorporation of said nucleotides or polynucleotides at said different defined positions.

13. The method of claim 10, wherein said nucleic acid array is an oligonucleotide array.

14. The method of claim 10, wherein the sequence of at least a portion of said target nucleic acid is identified.

15. The method of claim 10, wherein at least one of said nucleic acids associated with said array is single stranded.

* * * * *

EXHIBIT 2

US006403320B1

(12) **United States Patent**
Read et al.(10) **Patent No.: US 6,403,320 B1**
(45) **Date of Patent: *Jun. 11, 2002**(54) **SUPPORT BOUND PROBES AND METHODS OF ANALYSIS USING THE SAME**(75) Inventors: **J. Leighton Read; Stephen P. A. Fodor**, both of Palo Alto; **Lubert Stryer**, Stanford, all of CA (US); **Michael C. Pirrung**, Chapel Hill, NC (US)(73) Assignee: **Affymetrix, Inc.**, Santa Clara, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **09/684,377**(22) Filed: **Oct. 5, 2000****Related U.S. Application Data**

(63) Continuation of application No. 09/557,875, filed on Apr. 24, 2000, which is a continuation of application No. 09/056,927, filed on Apr. 8, 1998, now Pat. No. 6,197,506, which is a continuation of application No. 08/670,118, filed on Jun. 25, 1996, now Pat. No. 5,800,992, which is a division of application No. 08/168,904, filed on Dec. 15, 1993, now abandoned, which is a continuation of application No. 07/624,114, filed on Dec. 6, 1990, now abandoned, which is a continuation-in-part of application No. 07/492,462, filed on Mar. 7, 1990, now Pat. No. 5,143,854, and a continuation-in-part of application No. 07/362,901, filed on Jun. 7, 1989, now abandoned, application No. 09/684,377, which is a continuation-in-part of application No. 08/348,471, filed on Nov. 30, 1994, which is a continuation of application No. 07/805,727, filed on Dec. 6, 1991, now Pat. No. 5,424,186, which is a continuation-in-part of application No. 07/624,120, filed on Dec. 6, 1990, which is a continuation-in-part of application No. 07/492,462, filed on Mar. 7, 1990, now Pat. No. 5,143,854, which is a continuation-in-part of application No. 07/362,901, filed on Jun. 7, 1989, now abandoned.

(51) **Int. Cl.⁷** **C12Q 1/68**
(52) **U.S. Cl.** **435/6; 435/7.1**
(58) **Field of Search** **435/6, 91.2, 7.1**(56) **References Cited****U.S. PATENT DOCUMENTS**

3,730,844 A	5/1973	Gilham et al.	195/103.5 R
3,849,137 A	11/1974	Barzynski et al.	96/97
3,862,056 A	1/1975	Hartman	252/511
3,939,350 A	2/1976	Kronick et al.	250/365
4,072,576 A	2/1978	Arwin et al.	195/103.5 R
4,121,222 A	10/1978	Diebold et al.	347/7
4,180,739 A	12/1979	Abu-Shumays	250/461 R
4,216,245 A	8/1980	Johnson	427/2.13
4,238,757 A	12/1980	Schenck	357/25
4,269,933 A	5/1981	Pazos	430/291
4,314,821 A	2/1982	Rice	23/230 B
4,327,073 A	4/1982	Huang	424/1
4,339,528 A	7/1982	Goldman	430/323
4,342,905 A	8/1982	Fujii et al.	250/201
4,373,071 A	2/1983	Itakura	525/375
4,395,486 A	7/1983	Wilson et al.	435/6
4,405,771 A	9/1983	Jagur	528/266
4,444,878 A	4/1984	Paulus	435/7

4,444,892 A	4/1984	Malmros	436/528
4,448,534 A	5/1984	Wertz et al.	356/435
4,458,066 A	7/1984	Caruthers et al.	536/27
4,477,556 A	10/1984	Dueber et al.	430/281
4,478,967 A	10/1984	Eian et al.	524/86
4,483,920 A	11/1984	Gillespie et al.	435/6
4,500,707 A	2/1985	Caruthers et al.	536/27
4,500,919 A	2/1985	Schreiber	358/78
4,516,833 A	5/1985	Fusek	350/162.12
4,517,338 A	5/1985	Urdea et al.	525/54.11
4,533,682 A	8/1985	Tortorello et al.	523/414
4,537,861 A	8/1985	Elings et al.	436/518
4,542,102 A	9/1985	Dattagupta et al.	435/6
4,555,490 A	11/1985	Merril	436/86
4,556,643 A	12/1985	Paau et al.	435/5
4,562,157 A	12/1985	Lowe et al.	435/291
4,563,419 A	1/1986	Ranki et al.	435/6
4,569,967 A	2/1986	Kornreich et al.	525/54.11
4,580,895 A	4/1986	Patel	356/39
4,584,277 A	4/1986	Ullman	436/501
4,588,682 A	5/1986	Groet et al.	435/6
4,591,570 A	5/1986	Chang	435/7.24
4,598,049 A	7/1986	Zelinka et al.	422/116
4,613,566 A	9/1986	Potter	435/6
4,624,915 A	11/1986	Schindler et al.	435/4
4,626,684 A	12/1986	Landa	250/328
4,631,211 A	12/1986	Houghten	428/35
4,637,861 A	1/1987	Krull et al.	204/1 T
4,656,127 A	4/1987	Mundy	435/6

(List continued on next page.)

FOREIGN PATENT DOCUMENTS

CA	1284931	6/1991
DE	2242394	3/1974
DE	3440141	5/1986

(List continued on next page.)

OTHER PUBLICATIONSBallard et al., "Imaging Genes, Chromosomes and Nuclear Structures Using Laser-Scanning Confocal Microscopy," SPIE, *Bioimaging and Two-Dimensional Spectroscopy*, 1205:1-10, conference held Jan. 18-19, 1990, Los Angeles, CA., abstract also included (1990).Burns et al., "Scanning Silt Aperture Confocal Microscopy for Three-Dimensional Imaging," *Scanning*, 12:156-160 (1990).Frank et al., "Facile and rapid 'spot-synthesis' of large numbers of peptides on membrane sheets," *Proc. 21st European Pept. Symp.*, Platja D'Oro, Spain, Sep 2-8, 1990.

(List continued on next page.)

Primary Examiner—Eggerton A. Campbell(74) *Attorney, Agent, or Firm*—Philip L. McGarrigle; Joseph Liebschuetz; Lisa Treannie(57) **ABSTRACT**

The present invention provides methods and apparatus for sequencing, fingerprinting and mapping biological macromolecules, typically biological polymers. The methods make use of a plurality of sequence specific recognition reagents which can also be used for classification of biological samples, and to characterize their sources.

131 Claims, 27 Drawing Sheets

US 6,403,320 B1

Page 2

U.S. PATENT DOCUMENTS			
4,670,380 A	6/1987	Dattagupta	435/6
4,677,054 A	6/1987	White et al.	435/6
4,681,859 A	7/1987	Kramer	436/501
4,683,195 A	7/1987	Mullis et al.	435/6
4,683,202 A	7/1987	Mullis	435/91
4,689,405 A	8/1987	Frank et al.	536/27
4,704,353 A	11/1987	Humphries et al.	435/4
4,711,955 A	12/1987	Ward et al.	536/29
4,713,326 A	12/1987	Dattagupta et al.	435/6
4,713,347 A	12/1987	Mitchell et al.	436/501
4,715,413 A	12/1987	Backlund et al.	141/94
4,715,929 A	12/1987	Ogawa	156/643
4,716,106 A	12/1987	Chiswell	435/6
4,719,179 A	1/1988	Barany	435/172.1
4,719,615 A	1/1988	Feyrer et al.	369/284
4,722,906 A	2/1988	Guire	436/501
4,728,502 A	3/1988	Hamill	422/116
4,728,591 A	3/1988	Clark et al.	430/5
4,731,325 A	3/1988	Palva et al.	435/6
4,737,344 A	4/1988	Koizumi et al.	422/100
4,755,458 A	7/1988	Rabbani et al.	435/5
4,758,727 A	7/1988	Tomei et al.	250/458.1
4,762,881 A	8/1988	Kauer	525/54.11
4,766,062 A	8/1988	Diamond et al.	435/6
4,767,700 A	8/1988	Wallace	435/6
4,777,019 A	10/1988	Dandekar	422/68
4,780,504 A	10/1988	Buendia et al.	525/54.11
4,786,170 A	11/1988	Groeblor	356/318
4,786,684 A	11/1988	Glass	525/54.1
4,794,150 A	12/1988	Steel	525/54.11
4,808,508 A	2/1989	Platzer	430/143
4,810,869 A	3/1989	Yabe et al.	250/201
4,811,062 A	3/1989	Tabata et al.	356/152
4,811,218 A	3/1989	Hunkapiller et al.	204/461
4,812,512 A	3/1989	Buendia et al.	525/54.11
4,820,630 A	4/1989	Taub	435/5
4,822,566 A	4/1989	Newman	422/68
4,833,092 A	5/1989	Geysen	436/501
4,844,617 A	7/1989	Kelderman et al.	356/372
4,846,552 A	7/1989	Veldkamp et al.	350/162.2
4,849,513 A	7/1989	Smith et al.	536/27
4,855,225 A	8/1989	Fung et al.	435/6
4,865,990 A	9/1989	Stead et al.	435/803
4,868,103 A	9/1989	Stavrianopoulos et al.	435/5
4,874,500 A	10/1989	Madou et al.	204/412
4,877,745 A	10/1989	Hayes et al.	436/166
4,886,741 A	12/1989	Schwartz	435/5
4,888,278 A	12/1989	Singer et al.	435/6
4,921,805 A	5/1990	Gebeyehu et al.	435/270
4,923,901 A	5/1990	Koester et al.	521/53
4,925,785 A	5/1990	Wang et al.	435/6
4,931,384 A	6/1990	Layton et al.	435/7.31
4,946,942 A	8/1990	Fuller et al.	530/335
4,965,188 A	10/1990	Mullis et al.	435/6
4,973,493 A	11/1990	Guire	427/2
4,979,959 A	12/1990	Guire	623/66
4,981,783 A	1/1991	Augenlicht	435/6
4,981,985 A	1/1991	Kaplan et al.	556/50
4,984,100 A	1/1991	Takayama et al.	360/49
4,987,065 A	1/1991	Stavrianopoulos et al.	435/5
4,988,617 A	1/1991	Landegren et al.	435/6
4,992,383 A	2/1991	Farnsworth	436/89
4,994,373 A	2/1991	Stavrianopoulos et al.	435/6
5,002,867 A	3/1991	Macevicz	435/6
5,006,464 A	4/1991	Chu et al.	435/7.1
5,011,770 A	4/1991	Kung et al.	435/6
5,013,669 A	5/1991	Peters, Jr. et al.	436/518
5,021,550 A	6/1991	Zeiger	530/334
5,026,773 A	6/1991	Steel	525/54.11
5,026,840 A	6/1991	Dattagupta et al.	536/27
5,028,525 A	7/1991	Gray et al.	435/6
5,028,545 A	7/1991	Soini	436/501
5,037,882 A	8/1991	Steel	525/54.11
5,043,265 A	8/1991	Tanke et al.	435/6
5,047,524 A	9/1991	Andrus et al.	536/27
5,064,754 A	11/1991	Mills	435/6
5,077,085 A	12/1991	Schnur et al.	427/98
5,077,210 A	12/1991	Eigler et al.	435/176
5,079,600 A	1/1992	Schnur et al.	357/4
5,081,584 A	1/1992	Omichinski et al.	364/497
5,082,830 A	1/1992	Brakel et al.	514/44
5,091,652 A	2/1992	Mathies et al.	250/458.1
5,096,807 A	3/1992	Leaback	435/6
5,100,626 A	3/1992	Levin	422/100
5,100,777 A	3/1992	Chang	435/7.24
5,112,962 A	5/1992	Letsinger et al.	536/27
5,141,813 A	8/1992	Nelson	428/402
5,143,854 A	9/1992	Pirung et al.	436/518
5,149,625 A	9/1992	Church et al.	435/6
5,153,319 A	10/1992	Caruthers et al.	536/27
5,164,319 A	11/1992	Hafeman et al.	435/287.1
5,171,695 A	12/1992	Ekins	436/501
5,188,963 A	2/1993	Stapleton	435/288.3
5,192,980 A	3/1993	Dixon et al.	356/326
5,200,051 A	4/1993	Cozzette et al.	204/403
5,202,231 A	4/1993	Drmanac et al.	435/6
5,206,137 A	4/1993	Ip et al.	435/6
5,215,882 A	6/1993	Bahl et al.	435/6
5,215,889 A	6/1993	Schultz	435/41
5,219,726 A	6/1993	Evans	435/6
5,225,326 A	7/1993	Bresser et al.	435/6
5,232,829 A	8/1993	Longiaru et al.	435/6
5,235,028 A	8/1993	Barany et al.	528/335
5,242,974 A	9/1993	Holmes	525/54.11
5,252,743 A	10/1993	Barrett et al.	548/303.7
5,256,549 A	10/1993	Urdea et al.	435/91
5,258,506 A	11/1993	Urdea et al.	536/23.1
5,306,641 A	4/1994	Saccocio	436/85
5,310,893 A	5/1994	Erllich et al.	536/24.31
5,324,633 A	6/1994	Fodor et al.	435/6
5,328,824 A	7/1994	Ward et al.	435/6
5,348,855 A	9/1994	Dattagupta et al.	435/6
5,384,261 A	1/1995	Winkler et al.	436/518
5,405,783 A	4/1995	Pirung et al.	436/518
5,424,186 A	6/1995	Fodor et al.	435/6
5,424,188 A	6/1995	Schneider et al.	435/6
5,432,099 A	7/1995	Ekins	436/518
5,436,327 A	7/1995	Southern et al.	536/25.34
5,445,934 A	8/1995	Fodor et al.	435/6
5,447,841 A	9/1995	Gray et al.	435/6
5,474,796 A	12/1995	Brennan	427/2.13
5,486,452 A	1/1996	Gordon et al.	435/5
5,489,507 A	2/1996	Chehab	435/6
5,489,678 A	2/1996	Fodor et al.	536/22.1
5,492,806 A	2/1996	Drmanac et al.	435/5
5,494,810 A	2/1996	Barany et al.	435/91.52
5,510,270 A	4/1996	Fodor et al.	436/518
5,525,464 A	6/1996	Drmanac et al.	435/6
5,527,681 A	6/1996	Holmes	435/6
5,552,270 A	9/1996	Khrapko et al.	435/6
5,556,961 A	9/1996	Foote et al.	536/27.1
5,561,071 A	10/1996	Hollenberg et al.	437/1
5,569,584 A	10/1996	Augenlicht	435/6
5,571,639 A	11/1996	Hubbell et al.	430/5
5,593,839 A	1/1997	Hubbell et al.	435/6
5,599,720 A	2/1997	Ekins	436/501
5,604,099 A	2/1997	Erllich et al.	435/6
5,643,728 A	7/1997	Slater et al.	435/6
5,653,939 A	8/1997	Hollis et al.	422/50
5,667,667 A	9/1997	Southern	205/687
5,667,972 A	9/1997	Drmanac et al.	435/6

US 6,403,320 B1

Page 3

5,695,940	A	12/1997	Drmanac et al.	435/6	JP	63-223557	9/1988
5,698,393	A	12/1997	Macioszek et al.	435/5	JP	1-233447	9/1989
5,700,637	A	12/1997	Southern	435/6	NO	P 913186	8/1991
5,707,806	A	1/1998	Shuber	435/6	WO	WO 84/03151	8/1984
5,744,305	A	4/1998	Fodor et al.	435/6	WO	WO 84/03564	9/1984
5,776,737	A	7/1998	Dunn	435/91.1	WO	WO 85/01051	3/1985
5,777,888	A	7/1998	Rine et al.	364/496	WO	WO 86/00991	2/1986
5,800,992	A	9/1998	Fodor et al.	435/6	WO	WO 86/06487	11/1986
5,807,522	A	9/1998	Brown et al.	422/50	WO	WO 87/05942	10/1987
5,830,645	A	11/1998	Pinkel et al.	435/6	WO	WO 88/01058	2/1988
5,843,767	A	12/1998	Beattie	435/287.1	WO	WO 88/04777	6/1988
5,846,708	A	12/1998	Hollis et al.	435/6	WO	WO 89/05616	6/1989
5,869,237	A	2/1999	Ward et al.	435/6	WO	WO 89/08834	9/1989
5,871,697	A	2/1999	Rothberg et al.	422/68.1	WO	WO 89/10977	11/1989
5,972,619	A	10/1999	Drmanac et al.	435/6	WO	WO 89/11548	11/1989
6,018,041	A	1/2000	Drmanac et al.	536/24.3	WO	WO 89/12819	12/1989
6,025,136	A	2/2000	Drmanac et al.	435/6	WO	WO 90/00626	1/1990
6,040,166	A	3/2000	Erlich et al.	435/194	WO	WO 90/00887	2/1990
6,054,270	A	4/2000	Southern	435/6	WO	WO 90/15070	2/1990

FOREIGN PATENT DOCUMENTS

DE	3505287	3/1988	WO	WO 90/03789	5/1990
EP	046 083	2/1982	WO	WO 90/07582	7/1990
EP	088 636	9/1983	WO	WO 91/00868	1/1991
EP	103 197	3/1984	WO	WO 91/04266	4/1991
EP	127 438	12/1984	WO	WO 91/07087	5/1991
EP	063 810	3/1986	WO	WO 92/16655	1/1992
EP	174 879	3/1986	WO	WO 92/10092	6/1992
EP	194 132	9/1986	WO	WO 92/10588	6/1992
EP	228 075	7/1987	WO	WO 93/02992	2/1993
EP	233 403	8/1987	WO	WO 93/09668	5/1993
EP	245 662	11/1987	WO	WO 88/01302	6/1993
EP	268 237	5/1988	WO	WO 93/11262	6/1993
EP	130 523	6/1988	WO	WO 93/17126	9/1993
EP	281 927	9/1988	WO	WO 93/22456	11/1993
EP	228 310	10/1988	WO	WO 93/22480	11/1993
EP	288 310	10/1988	WO	WO 95/00530	1/1995
EP	304 202	2/1989	WO	WO 95/11995	5/1995
EP	307 476	3/1989	WO	WO 95/33846	12/1995
EP	319 012	6/1989	WO	WO 96/23078	8/1996
EP	328 256	8/1989	WO	WO 97/10365	3/1997
EP	333 561	9/1989	WO	WO 97/17317	5/1997
EP	337 498	10/1989	WO	WO 97/19410	5/1997
EP	386 229	4/1990	WO	WO 97/27317	7/1997
EP	373 203	6/1990	WO	WO 97/29212	8/1997
EP	392 546	10/1990	WO	WO 97/31256	8/1997
EP	142 299	12/1990	WO	WO 97/45559	12/1997
EP	400 920	12/1990	WO	WO 98/03673	1/1998
EP	173 339	1/1992	WO	WO 98/31836	7/1998
EP	171 150	3/1992	YU	P-570/87	4/1987
EP	237 362	3/1992	YU	18617/87	9/1987
EP	185 547	6/1992	OTHER PUBLICATIONS		
EP	260 634	6/1992			
EP	232 967	4/1993	Southern et al., "Parallel synthesis and analysis of large numbers of related chemical compounds: applications to oligonucleotides," <i>J. Biotechnology</i> , 35:217-227 (1994).		
EP	235 726	5/1993	Burns et al., "Scanning Slit Aperture Confocal Microscopy for Three-Dimensional Imaging," <i>Scanning</i> , vol. 12, 156-160 (1990).		
EP	476 014	8/1994	Ballard, S., "Imaging Genes, Chromosomes and Nuclear Structures using Laser-Scanning Confocal Microscopy," <i>SPIE</i> , vol. 1205, 2-10 (1990).		
EP	225 807	10/1994	Sequencing by Hybridization Workshop, listing of participants and workshop presentation summaries, from workshop held Nov. 19-20, 1991.		
EP	717 113	6/1996	"A Sequencing Reality Check," <i>Science</i> , 242:1245 (1988).		
EP	721 016	7/1996	"Affymax raises \$25 million to develop high-speed drug discovery system," <i>Biotechnology News</i> , 10(3):7-8 (1990).		
EP	535 242	9/1997			
EP	848 067	6/1998			
FR	619 321	1/1999			
FR	2559783	3/1988			
GB	2156074	3/1988			
GB	2196476	4/1988			
GB	8810400.5	5/1988			
GB	2233654	1/1991			
GB	2248840	9/1992			
JP	49-110601	10/1974			
JP	60-248669	12/1985			
JP	63-084499	4/1988			

US 6,403,320 B1

Page 4

- "Preparation of fluorescent-labeled DNA and its use as a probe in molecular hybridization," *Bioorg Khim*, 12(11):1508-1513 (1986).
- Abbott et al., "Manipulation of the Wettability of Surfaces on the 0.1-to 1-Micrometer Scale Through Micromachining and Molecular Self-Assembly," *Science*, 257:1380-1382 (1992).
- Adams et al., "Complementary DNA Sequencing: Expressed Sequence Tags and Human Genome Project," *Science*, 252(5013):1651-1656 (1991).
- Adams et al., "Photolabile Chelators That "Cage" Calcium with Improved Speed of Release and Pre-Photolysis Affinity," *J. Gen. Physiol.*, p. 9a (12/86).
- Adams et al., "Biologically Useful Chelators That Take Up Ca^{2+} upon Illumination," *J. Am. Chem. Soc.*, 111:7957-7968 (1989).
- Ajayaghosh et al., "Solid-Phase Synthesis of N-Methyl- and N-Ethylamides of Peptides Using Photolytically Detachable ((3-Nitro-4((alkylamino)methyl)benzamido)methyl)polystyrene Resin," *J. Org. Chem.*, 55(9):2826-2829 (1990).
- Ajayaghosh et al., "Solid-phase synthesis of C-terminal peptide amides using a photoremovable α -methylphenacylamido anchoring linkage," *Proc. Ind. Natl. Sci. (Chem.Sci.)*, 100(5):389-396 (1988).
- Ajayaghosh et al., "Polymer-supported Solid-phase Synthesis of C-Terminal Peptide N-Methylamides Using a Modified Photoremovable 3-Nitro-4-N-methylaminomethylpolystyrene Support," *Ind. J. Chem.*, 27B:1004-1008 (1988).
- Ajayaghosh et al., "Polymer-Supported Synthesis of Protected Peptide Segments on a Photosensitive o-Nitro(α -Methyl)Bromobenzyl Resin," *Tetrahedron*, 44(21):6661-6666 (1988).
- Amit et al., "Photosensitive Protecting Groups of Amino Sugars and Their Use in Glycoside Synthesis. 2-Nitrobenzyloxycarbonylamino and 6-Nitroveratryloxycarbonylamino Derivatives," *J. Org. Chem.*, 39(2):192-196 (1974).
- Amit et al., "Photosensitive Protecting Groups—A Review," *Israel J. Chem.*, 12(1-2):103-113 (1974).
- Anand et al., "A 3.5 genome equivalent multi access YAC library: construction, characterisation, screening and storage," *Nuc. Acids Res.*, 18(8):1951-1956 (1990).
- Anderson et al., "Quantitative Filter Hybridisation," chapter 3 from *Nucleic Acid Hybridization a practical approach*, pp. 73-111, Hames et al., eds., IRL Press (1985).
- Applied Biosystems. Model 431A Peptide Synthesizer User's manual, Sections 2 and 6, (Aug. 15, 1989).
- Arnold et al., "A Novel Universal Support for DNA & RNA Synthesis," abstract from *Federation Proceedings*, 43(7):abstract No. 3669 (1984).
- Atherton et al., *Solid Phase Peptide Synthesis: A Practical Approach*, IRL Press, (1989), tbl. of cont., pp. vii-ix.
- Augenlicht et al., "Cloning and Screening of Sequences Expressed in a Mouse Colon Tumor," *Cancer Research*, 42:1088-1093 (1982).
- Augenlicht et al., "Expression of Cloned Sequences in Biopsies of Human Colonic Tissue and in Colonic Carcinoma Cells Induced to Differentiate in Vitro," *Cancer Res.*, 47:6017-6021 (1987).
- Bains, W., "Hybridization Methods for DNA Sequencing," *Genomics*, 11(2):294-301 (1991).
- Bains et al., "A Novel Method for Nucleic Acid Sequence Determination," *J. Theor. Biol.*, 135:303-307 (1988).
- Bains, W., "Alternative Routes Through the Genome," *Bio-technology*, 8:1251-1256 (1988).
- Balachander et al., "Functionalized Siloxy-Anchored Monolayers with Exposed Amino, Azido, Bromo, or Cyano Groups," *Tetrahed. Ltrs.*, 29(44):5593-5594 (1988).
- Baldwin et al., "New Photolabile Phosphate Protecting Groups," *Tetrahed.*, 46(19):6879-6884 (1990).
- Bannwarth et al., "Laboratory Methods, A System for the Simultaneous Chemical synthesis of Different DNA Fragments on Solid Support," *DNA*, 5(5):413-419 (1986).
- Bannwarth, W., "Gene Technology: a Challenge for a Chemist," *CHIMIA*, 41(9):302-317 (1987).
- Barany, F., "Genetic disease detection and DNA amplification using cloned thermostable ligase," *PNAS*, 88:189-193 (1991).
- Barltrop et al., "Photosensitive Protective Groups," *Chemical Communications*, pp. 822-823 (1966).
- Barinaga, M., "Will 'DNA Chip' Speed Genome Initiative," *Science*, 253:1489 (1985).
- Bart et al., "Microfabricated Electrohydrodynamic Pumps," *Sensors and Actuators*, A21-A23:193-197 (1990).
- Bartsh et al., "Cloning of mRNA sequences from the human colon: Preliminary characterisation of defined mRNAs in normal and neoplastic tissues," *Br. J. Can.*, 54:791-798 (1986).
- Baum, R., "Fledgling firm targets drug discovery process," *Chem. Eng. News*, p. 10-11 (1990).
- Beltz et al., "Isolation of Multigene Families and Determination of Homologies by Filter Hybridization Methods," *Methods in Enzymology*, 100:266-285 (1983).
- Benschop, Chem. Abstracts 114(26):256643 (1991).
- Bhatia et al., "New Approach To Producing Patterned Biomolecular Assemblies," *J. American Chemical Society*, 114:4432-4433 (1992).
- Biorad Chromatography Electrophoresis Immunochemistry Molecular Biology HPLC catalog M 1987 pp. 182.
- Blawas et al., "Step-and-Repeat Photopatterning of Protein Features Using Caged-Biotin-BSA: Characterization and Resolution," *Langmuir*, 14(15):4243-4250 (1998).
- Blawas, A.S., "Photopatterning of Protein Features using Caged-biotin-Bovine Serum Albumin," dissertation for Ph.D at Duke University in 1998.
- Bos et al., "Amino-acid substitutions at codon 13 of the N-ras oncogene in human acute myeloid leukaemia," *Nature*, 315:726-730 (1985).
- Boyle et al., "Differential distribution of long and short interspersed element sequences in the mouse genome: Chromosome karyotyping by fluorescence in situ hybridization," *PNAS*, 87:7757-7761 (1990).
- Brock et al., "Rapid fluorescence detection of in situ hybridization with biotinylated bovine herpesvirus-1 DNA probes," *J. Veterinary Diagnostic Invest.*, 1:34-38 (1989).
- Burgi et al., "Optimization in Sample Stacking for High-Performance Capillary Electrophoresis," *Anal. Chem.*, 63:2042-2047 (1991).
- Cameron et al., "Photogeneration of Organic Bases from o-Nitrobenzyl-Derived Carbamates," *J. Am. Chem. Soc.*, 113:4303-4313 (1991).
- Carrano et al., "A High-Resolution, Fluorescence-Based, Semiautomated Method for DNA Fingerprinting," *Genomics*, 4:129-136 (1989).
- Caruthers, M.H., "Gene Synthesis Machines: DNA Chemistry and Its Uses," *Science*, 230:281-285 (1985).

US 6,403,320 B1

Page 5

- Chatterjee et al., "Inducible Alkylation of DNA Using an Oligonucleotide-Quinone Conjugate," *Am. J. Chem. Soc.*, 112:6397-6399 (1990).
- Chee et al., "Accessing Genetic Information with High-Density DNA Arrays," *Science*, 274:610-614 (1996).
- Chehab et al., "Detection of sickle cell anaemia mutation by colour DNA amplification," *Lancet*, 335:15-17 (1990).
- Chehab et al., "Detection of specific DNA sequences by fluorescence amplification: A color complementation assay," *PNAS*, 86:9178-9182 (1989).
- Chetverin et al., "Oligonucleotide Arrays: New Concepts and Possibilities," *Biotechnology*, 12:1093-1099 (1994).
- Church et al., "Multiplex DNA sequencing," *Science*, 240:185-188 (1988).
- Church et al., "Genomic sequencing," *PNAS*, 81:1991-1995 (1984).
- Clevite Corp., Piezoelectric Technology, Data for Engineers.
- Corbett et al., "Reaction of Nitroso Aromatics with Glyoxylic Acid. A New Path to Hydroxamic Acids," *J. Org. Chem.*, 45:2834-2839 (1980).
- Coulsen et al., "Toward a physical map of the genome of the nematode *Caenorhabditis elegans*," *PNAS*, 83:7821-7825 (1986).
- Craig et al., "Ordering of cosmid clones covering the Herpes simplex virus type 1 (HSV-1) genome: a test case for fingerprinting by hybridization," *Nuc. Acid. Res.*, 18(9):2653-2660 (1990).
- Cummings et al., "Photoactivable Fluorophores. 1. Synthesis and Photoactivation of o-Nitrobenzyl-Quenched Fluorescent Carbamates," *Tetrahedron Letters*, 29(1):65-68 (1988).
- Dattagupta et al., "Rapid identification of Microorganisms by Nucleic Acid Hybridization after Labeling the Test Sample," *Anal. Biochem.*, 177:85-89 (1989).
- Dattagupta et al., "Nucleic Acid Hybridization: a Rapid Method for the Diagnosis of Infectious Diseases," *Perspectives in Antiinfective Therapy*, eds. Jackson et al., pp. 241-247 (1988).
- Dower et al., "The Search for Molecular Diversity (II): Recombinant and Synthetic Randomized Peptide Libraries," *Ann. Rep. Med. Chem.*, 26:271-280 (1991).
- Diggelmann, "Investigating the VLSIPS synthesis process," Sep. 9, 1994.
- Di Mauro et al., "DNA Technology in Chip Construction," *Adv. Mater.*, 5(5):384-386 (1993).
- Drmanac et al., "An Algorithm for the DNA Sequence Generation from k-Tuple Word Contents of the Minimal Number of Random Fragments," *J. Biomol. Struct. Dyn.*, 8(5):1085-1102 (1991).
- Drmanac et al., "Partial Sequencing by Oligo-Hybridization Concept and Applications in Genome Analysis," 1st Int. Conf. Electrophor., Supercomp., Hum. Genome pp. 60-74 (1990).
- Drmanac et al., "Sequencing by Oligonucleotide Hybridization: A Promising Framework in Decoding of the Genome Program?," 1st Int. Conf. Electrophor., Supercomp., Hum. Genome pp. 47-59 (1990).
- Drmanac et al., "Laboratory Methods, Reliable Hybridization of Oligonucleotides as Short as Six Nucleotides," *DNA and Cell Biol.*, 9(7):527-534 (1990).
- Drmanac et al., "Sequencing of Megabase Plus DNA by Hybridization: theory of the Method," *Genomics*, 4:114-128 (1989).
- Dramanac et al., "Sequencing of Megabase Plus DNA by Hybridization: Theory of the Method," abstract of presentation given at Cold Spring Harbor Symposium on Genome Mapping and Sequencing, Apr. 27, 1988 thru May 1, 1988.
- Dulcey et al., "Deep UV Photochemistry of Chemisorbed Monolayers: Patterned Coplanar Molecular Assemblies," *Science*, 252:551-554 (1991).
- Duncan et al., "Affinity Chromatography of a Sequence-Specific DNA Binding Protein Using Teflon-Linked Oligonucleotides," *Analytical Biochemistry*, 169:104-108 (1988).
- Effenhauser et al., "Glass Chips for High-speed Capillary Electrophoresis Separations with Submicrometer Plate Heights," *Anal. Chem.*, 65:2637-2642 (1993).
- Effenhauser et al., "High-Speed Separation of Antisense Oligonucleotides on a Micromachined Capillary Electrophoresis Device," *Anal. Chem.*, 66:2949-2953 (1994).
- Ekins et al., "High Specific Activity Chemiluminescent and Fluorescent Markers: their Potential Application to High Sensitivity and 'Multi-analyte' Immunoassays," *J. Bioluminescence Chemiluminescence*, 4:59-78 (1989).
- Ekins et al., "Development of Microspot Multi-Analyte Ratiometric Immunoassay Using dual Fluorescent-Labelled Antibodies," *Anal. Chimica Acta*, 227:73-96 (1989).
- Ekins et al., "Multianalyte Microspot Immunoassay-Microanalytical 'Compact Disk' of the Future," *Clin. Chem.*, 37(11):1955-1967 (1991).
- Ekins, R.P., "Multi-analyte immunoassay*," *J. Pharmaceut. Biomedical Analysis*, 7(2):155-168 (1989).
- Ekins et al., "Fluorescence Spectroscopy and its Application to a New Generation of High Sensitivity, Multi-Microspot, Multianalyte, Immunoassay," *Clin. Chim. Acta*, 194:91-114 (1990).
- Elder, J.K., "Analysis of DNA Oligonucleotide Hybridization Data by Maximum Entropy," in *Maximum Entropy and Bayesian Methods*, eds. Mohammad-Djafari and Demoment, Kluwer, Dordrecht, pp. 363-371 (1992).
- Ellis, R.W., "The Application of Synthetic Oligonucleotides to Molecular Biology," *Pharmaceutical Research*, 3(4):195-207 (1986).
- Evans et al., "Microfabrication for Automation of Molecular processes in Human Genome Analysis," *Clin. Chem.*, 41(11):1681 (1995).
- Evans et al., "Physical mapping of complex genomes by cosmid multiplex analysis," *PNAS*, 86:5030-5034 (1989).
- Ezaki et al., "Small-Scale DNA Preparation for Rapid Genetic Identification of *Campylobacter* Species without Radioisotope," *Microbiol. Immunology*, 32(2):141-150 (1988).
- Fan et al., "Mapping small DNA sequences by fluorescence in situ hybridization directly on banded metaphase chromosomes," *PNAS*, 87(16):6223-6227 (1990).
- Fan et al., "Micromachining of Capillary Electrophoresis Injectors and Separators on Glass Chips and Evaluation of Flow at Capillary Intersections," *Anal. Chem.*, 66:177-184 (1994).
- Feinberg et al., Addendum to "A technique for Radiolabeling DNA Restriction Endonuclease Fragments to High Specific Activity," *Anal. Biochem.*, 137:266-267 (1984).
- Fettingner et al., "Stacked modules for micro flow systems in chemical analysis: concept and studies using an enlarged model," *Sensors and Actuators*, B17:19-25 (1993).
- Flanders et al., "A new interferometric alignment technique," *App. Phys. Lett.*, 31(7):426-429 (1977).

US 6,403,320 B1

Page 6

- Fodor et al., "Multiplexed biochemical assays with biological chips," *Nature*, 364:555-556 (1993).
- Fodor et al., "Light-directed, Spatially Addressable Parallel Chemical Synthesis," *Science*, 251:767-773 (1991).
- Forman et al., "Thermodynamics of Duplex Formation and Mismatch Discrimination on Photolithographically Synthesized Oligonucleotide Arrays," chapter 13pp. 206-228 from *Molecular Modeling of Nucleic Acids*, ACS Symposium Series 682, 4/13-17/97, Leontis et al., eds.
- Frank et al., "Simultaneous Multiple Peptide Synthesis Under Continuous flow Conditions on Cellulose Paper Discs as Segmental Solid Supports," *Tetrahedron*, 44(19):6031-6040 (1988).
- Frank et al., "Automation of DNA Sequencing Reactions and Related Techniques: A Workstation for Micromanipulation of Liquids," *BioTechnology*, 6:1211-1212 (1988).
- Frank et al., "Simultaneous Synthesis and Biological Applications of DNA Fragments: An Efficient and Complete Methodology," *Methods in Enzymology*, 154:221-250 (1987).
- Fuhr et al., "Traveling wave-driven microfabricated electrohydrodynamic pumps for liquids," *J. Micromech. Microeng.*, 4:217-226 (1994).
- Fuller et al., "Urethane-Protected Amino Acid N-Carboxy Anhydrides and Their Use in Peptide Synthesis," *J. Amer. Chem. Soc.*, 112(20):7414-7416 (1990).
- Furka et al., "General method for rapid synthesis of multi-component peptide mixtures," *Int. J. Peptide Protein Res.*, 37:487-493 (1991).
- Furka et al., "Cornucopia of Peptides by Synthesis," 14th Int. Congress of Biochem. abst#FR:013, 7/10-15/88 Prague, Czechoslovakia.
- Furka et al., "More Peptides by Less Labour," abst. 288, Int. Symp. Med. Chem., Budapest Hungary 8/15-19/88.
- Gait, eds., pp. 1-115 from *Oligonucleotide Synthesis: A Practical Approach*, IRL Press, (1984).
- Gazard et al., "Lithographic Technique Using Radiation-Induced Grafting of Acrylic Acid into Poly(Methyl Methacrylate) Films," *Polymer Engineering and Science*, 20(16):1069-1072 (1980).
- Gergen et al., "Filter replicas and permanent collections of recombinant DNA plasmids," *Nuc. Acids Res.*, 7(8):2115-2137 (1979).
- Getzoff et al., "Mechanisms of Antibody Binding to a Protein," *Science*, 235:1191-1196 (1987).
- Geysen et al., "Strategies for epitope analysis using peptide synthesis," *J. Immunol. Meth.*, 102:259-274 (1987).
- Geysen et al., "Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid," *PNAS*, 81:3998-4002 (1984).
- Geysen et al., "A synthetic strategy for epitope mapping," from *Peptides: Chem. & Biol.*, Proc. of 10th Am. Peptide Symp., 5/23-28/87, pp. 519-523, (1987).
- Geysen, "Antigen-antibody interactions at the molecular level: adventures in peptide synthesis," *Immunol. Today*, 6(12):364-369 (1985).
- Geysen et al., "Cognitive Features of Continuous Antigenic Determinants," from *Synthetic Peptides: Approaches to Biological Probes*, pp. 19-30, (1989).
- Geysen et al., "Chemistry of Antibody Binding to a Protein," *Science*, 235:1184-1190 (1987).
- Geysen et al., "The delineation of peptides able to mimic assembled epitopes," 1986 CIBA Symp., pp. 130-149.
- Geysen et al., "Cognitive Features of Continuous Antigenic Determinants," *Mol. Recognit.*, 1(1):1-10 (1988).
- Geysen et al., "A Prio Ri Delineation of a Peptide Which Mimics A Discontinuous Antigenic Determinant," *Mol. Immunol.*, 23(7):709-715 (1986).
- Ghosh et al., "Covalent attachment of oligonucleotides to solid supports," *Nuc. Acids Res.*, 15(13):5353-5373 (1987).
- Gilon et al., "Backbone Cyclization: A New Method for Conferring Conformational Constraint on Peptides," *Biopolymers*, 31(6):745-750 (1991).
- Gingeras et al., "Hybridization properties of immobilized nucleic acids," *Nuc. Acids Res.*, 15(13):5373-5390 (87).
- Gummerlock et al., "RAS Enzyme-Linked Immunoblot Assay Discriminates p21 Species: A Technique to Dissect Gene Family Expression," *Anal. Biochem.*, 180:158-168 (1989).
- Gurney et al., "Activation of potassium current by rapid photochemically generated step increases of intracellular calcium in rat sympathetic neurons," *PNAS*, 84:3496-3500 (1987).
- Haase et al., "Detection of Two Viral Genomes in Single Cells by Double-Label Hybridization in Situ and Color Microradioautography," *Science*, 227:189-192 (1985).
- Hacia, et al., "Two color hybridization analysis using high density oligonucleotide arrays and energy transfer dyes," *Nuc. Acids Res.*, 26(16):3865-3866 (1998).
- Hack, M.L., "Conics Formed to Make Fluid & Industrial Gas Micromachines," *Genetic Engineering News*, 15(18):1, 29 (1995).
- Hagedorn et al., "Pumping of Water Solutions in Microfabricated Electrohydrodynamic Systems," from *Micro Electro Mechanical Systems* conference in Travemunde Germany (1992).
- Hames et al., *Nuclear acid hybridization, a practical approach*, cover page and table of contents (1985).
- Hanahan et al., "Plasmid Screening at High Colony Density," *Meth. Enzymology*, 100:333-342 (1983).
- Hanahan et al., "Plasmid screening at high colony density," *Gene*, 10:63-67 (1980).
- Haridasan et al., "Peptide Synthesis using Photolytically Cleavable 2-Nitrobenzyloxycarbonyl Protecting Group," *Proc. Indian Natn. Sci. Acad.*, 53A(6):717-728 (1987).
- Harrison et al., "Capillary Electrophoresis and Sample Injection Systems Integrated on a Planar Glass Chip," *Anal. Chem.*, 64:1926-1932 (1992).
- Harrison et al., "Micromachining a Miniatured Capillary Electrophoresis-Based Chemical Analysis System on a Chip," *Science*, 261:895-897 (1993).
- Harrison et al., "Towards minaturized electrophoresis and chemical analysis systems on silicon: an alternative to chemical sensors*," *Sensors and Actuators*, B10:107-116 (1993).
- Harrison et al., "Rapid separation of fluorescein derivatives using a micromachined capillary electrophoresis system," *Analytica Chimica Acta*, 283:361-366 (1993).
- Hellberg et al., "Minimum analogue peptide sets (MAPS) for quantitative structure-activity relationships," *Int. J. Peptide Res.*, 37:414-424 (1991).
- Hilser et al., "Protein and peptide mobility in capillary zone electrophoresis, A comparison of existing models and further analysis," *J. Chromatography*, 630:329-336 (1993).
- Ho et al., "Highly Stable Biosensor Using an Artificial Enzyme," *Anal. Chem.*, 59:536-537 (1987).

US 6,403,320 B1

Page 7

- Hochgeschwender et al., "Preferential expression of a defined T-cell receptor β -chain gene in hapten-specific cytotoxic T-cell clones," *Nature*, 322:376-378 (1986).
- Hodgson, J., "Assays A La Photolithography," *Biotech.*, 9:419 (1991).
- Hodgson et al., "Hybridization probe size control: optimized 'oligolabelling'," *Nuc. Acids Res.*, 15(15):6295 (1987).
- Hopman et al., "Bi-color detection of two target DNAs by non-radioactive in situ hybridization*," *Histochem.*, 85:1-4 (1986).
- Iwamura et al., "1-Pyrenylmethyl Esters, Photolabile Protecting Groups for Carboxylic Acids," *Tetrahedron Ltrs.*, 28(6):679-682 (1987).
- Iwamura et al., "1-(α -Diazobenzyl)pyrene: A Reagent for Photolabile and Fluorescent Protection of Carboxyl Groups of Amino Acids and Peptides," *Synlett*, p. 35-36 (1991).
- Jacobson et al., "Effects of Injection Schemes and Column Geometry on the Performance of Microchip Electrophoresis Devices," *Anal. Chem.*, 66:1107-1113 (1994).
- Jacobsen et al., "Open Channel Electrochromatography on a Microchip," *Anal. chem.*, 66:2369-2373 (1994).
- Jacobson et al., "Microchip Capillary Electrophoresis with an Integrated Postcolumn Reactor" *Anal. Chem.*, 66:3472-3476 (1994).
- Jacobson et al., "Precolumn Reactions with Electrophoretic Analysis Integrated on a Microchip," *Anal. Chem.*, 66:4127-4132 (1994).
- Jacobson et al., "Microfabricated chemical measurement systems," *Nature Medicine*, 1(10):1093-1096 (1995).
- Jacobsen et al., "Fused Quartz Substrates for Microchip Electrophoresis," *Anal. chem.*, 67:2059-2063 (1995).
- Jacobson et al., "High-Speed Separations on a Microchip," *Anal. Chem.*, 66:1114-1118 (1994).
- Jacobson et al., "Microchip electrophoresis with sample stacking," *Electrophoresis*, 16:481-486 (1995).
- Jayakumari, "Peptide synthesis in a triphasic medium catalysed by papain immobilized on a crosslinked polystyrene support," *Indian J. Chemistry*, 29B:514-517 (1990).
- Jovin et al., "Luminescence Digital Imaging Microscopy," *Ann. Rev. Biophys. Biophys. Chem.*, 18:271-308 (1989).
- Kafatos et al., "Determination of nucleic acid sequence homologies and relative concentrations by a dot hybridization procedure," *Nuc. Acids Res.*, 7(6):1541-1553 (1979).
- Kaiser et al., "Peptide and Protein Synthesis by Segment Synthesis-Condensation," *Science*, 243:187-192 (1989).
- Kaplan et al., "Photolabile chelators for the rapid photorelease of divalent cations," *PNAS*, 85:6571-6575 (1988).
- Karube, "Micro-biosensors based on silicon fabrication technology," chapter 25 from *Biosensors: Fundamentals and Applications*, Turner et al., eds., Oxford Publ., 1987, pp. 471-480 (1987).
- Kates et al., "A Novel, Convenient, Three-dimensional Orthogonal Strategy for Solid-Phase Synthesis of Cyclic Peptides 1-3," *Tetrahed. Letters*, 34(10):1549-1552 (1993).
- Kerkof et al., "A Procedure for Making Simultaneous Determinations of the Relative Levels of Gene Transcripts in Tissues of Cells," *Anal. Biochem.*, 188:349-355 (1990).
- Khrapko et al., "An Oligonucleotide hybridization approach to DNA sequencing," *FEBS Lett.*, 256(1,2):118-122 (1989).
- Khrapko et al., "A method for DNA sequencing by hybridization with oligonucleotide matrix," *DNA Seq. Map.*, 1:375-388 (1991).
- Kidd et al., " α_1 -Antitrypsin deficiency detection by direct analysis of the mutation in the gene," *Nature*, 304:230-234 (1983).
- Kievits et al., "Rapid subchromosomal localization of cosmid by nonradioactive in situ hybridization," *Cytogenetics Cell Genetics*, 53(2-3):134-136 (1990).
- Kimura et al., "An Immobilized Enzyme Membrane Fabrication Method using an Ink Jet Nozzle," *Biosensors*, 4:41-52 (1988).
- Kimura et al., "An Integrated SOS/FET Multi-Biosensor," *Sensors & Actuators*, 9:373-387 (1986).
- Kitazawa et al., "In situ DNA-RNA hybridization using in vivo bromodeoxyuridine-labeled DNA probe," *Histochemistry*, 92:195-199 (1989).
- Kleinfeld et al., "Controlled Outgrowth of Dissociated Neurons on Patterned Substrates," *J. Neurosci.*, 8(11):4098-4120 (1988).
- Knight, P., "Materials and Methods/Microsequencers for Proteins and Oligosaccharides," *Bio/Tech.*, 7:1075-76 (1989).
- Kohara et al., "The Physical Map of the Whole *E. coli* Chromosome: Application of a New Strategy for Rapid Analysis and Sorting of a Large Genomic Library," *Cell*, 50:495-508 (1987).
- Krile et al., "Multiplex holography with chirp-modulated binary phase-coded reference-beam masks," *Applied Opt.*, 18(1):52-56 (1979).
- Labat, I., "Subfragments as an informative characteristic of the DNA molecule—computer simulation," research report submitted to the University of Belgrade College of Natural Sciences and Mathematics, (1988).
- Lander et al., "Genomic Mapping by Fingerprinting Random Clones: A Mathematical Analysis," *Genomics*, 2:231-239 (1988).
- Lainer et al., "Human Lymphocyte Subpopulations Identified by Using Three-Color Immunofluorescence and Flow Cytometry of Analysis: Correlation of Leu-2, Leu-3, Leu-7, Leu-8, and Leu-11 Clee Surface Antigen Expression," *Journal of Immunology*, 132(1):151-156 (1984).
- Lam et al., "A new type of synthetic peptide library for identifying ligand-binding activity," *Nature*, 354:82-84 (1991).
- Laskey et al., "Messenger RNA prevalence in sea urchin embryos measured with cloned cDNAs," *PNAS*, 77(9):5317-5321 (1980).
- Lee et al., "synthesis of a Polymer Surface Containing Covalently Attached Triethoxysilane Functionality Adhesion to Glass," *Macromolecules*, 21:3353-3356 (1988).
- Lehrach et al., "Labeling oligonucleotides to high specific activity(I)," *Nuc. Acids Res.*, 17(12):4605-4610 (89).
- Lehrach et al., "Phage Vectors—EMBL Series," *Meth. Enzymology*, 153:103-115 (1987).
- Lehrach et al., "Hybridization Fingerprinting in Genome Mapping and Sequencing," *Genome Analysis vol. 1: Genetic and Physical Mapping*, Cold Spring Harbor Laboratory Press, pp. 39-81 (1990).
- Levy, M.F., "Preparing Additive Printed Circuits," *IBM Tech. Discl. Bull.*, 9(11):1473 (1967).
- Lewin, Benjamin, eds., *Genes*, third edition, John Wiley & Sons, cover page, preface and table of contents, (1987).
- Lichter et al., "High-Resolution Mapping of Human Chromosome 11 by in Situ hybridization with Cosmid Clones," *Science*, 247:64-69 (1990).

US 6,403,320 B1

Page 8

- Lichter et al., "Fluorescence in situ hybridization with Alu and L1 polymerase chain reaction probes for rapid characterization of human chromosomes in hybrid cell lines," *PNAS*, 87:6634-6638 (1990).
- Lichter et al., "Rapid detection of human chromosome 21 aberrations by in situ hybridization," *PNAS*, 85:9664-9668 (1988).
- Lichter et al., "Is non-isotopic in situ hybridization finally coming of age," *Nature*, 345:93-94 (1990).
- Lieberman et al., "A Light source Smaller Than the Optical Wavelength," *Science*, 247:59-61 (1990).
- Lipshutz et al., "U sing Oligonucleotide Probe Arrays To Access Genetic Diversity," *BioTech.*, 19(3):442-7 (1995).
- Little, P., "Clone maps made simple," *Nature*, 346:611-612 (1990).
- Liu et al., "Sequential Injection Analysis in Capillary Format with an Electroosmotic Pump," *Talanta*, 41(11):1903-1910 (1994).
- Lockhart et al., "Expression monitoring by hybridization to high-density oligonucleotide arrays," *Nat. Biotech.*, 14:1675-1680 (1996).
- Logue et al., "General Approaches to Mask Design for Binary Optics," *SPIE*, 1052:19-24 (1989).
- Loken et al., "three-color Immunofluorescence Analysis of Leu Antigens on Human Peripheral Blood Using Two Lasers on a Fluorescence-Activated Cell Sorter," *Cytoetry*, 5:151-158 (1984).
- Love et al., "Screening of λ Library for Differentially Expressed Genes Using in Vitro Transcripts," *Anal. Biochem.*, 150:429-441 (1985).
- Lowe, C.R., "Biosensors," *Trends in Biotech.*, 2:59-65 (1984).
- Lowe, C.R., "An Introduction to the Concepts and Technology of Biosensors," *Biosensors*, 1:3-16 (1985).
- Lowe, C.R., *Biotechnology and Crop Improvement and Protection*, BCPC Publications, pp. 131-138 (1986).
- Lowe et al., "Solid-Phase Optoelectronic Biosensors," *Methods in Enzymology*, 137:338-347 (1988).
- Lowe, C.R., "Biosensors," *Phil. Tran. R. Soc. Lond.*, 324:487-496 (1989).
- Lu et al., "Differential screening of murine ascites cDNA libraries by means of in vitro transcripts of cell-cycle-phase-specific cDNA and digital image processing," *Gene*, 86:185-192 (1990).
- Luo, J. et al., "Improving the fidelity of *Thermus thermophilus* DNA ligase," *Nuc. Acids Res.*, 24(14):3071-3078 (1996).
- Lysov et al., "A new method for determining the DNA nucleotide sequence by hybridization with oligonucleotides," *Doklady Biochem.*, 303(1-6):436-438 (1989).
- Lysov et al., "DNA Sequencing by Oligonucleotide Hybridization," First International Conference on Electrophoresis, Supercomputing and the Human Genome, 4/10-13/90 p. 157.
- MacDonald et al., "A Rapid ELISA for Measuring Insulin in Large Number of Research Samples," *Metabolism*, 38(5):450-452 (1989).
- Mairanovsky, V.G., "Electron-Deprotection-Electrochemical Removal of Protecting Groups**," *Agnew. Chem. Int. Ed. Engl.*, 15(5):281-292 (1976).
- Manz et al., "Miniaturized Total Chemical Analysis Systems: a Novel Concept for Chemical Sensing," *Sensors and Actuators*, B1:244-248 (1990).
- Manz et al., "Micromachining of monocrystalline silicon and glass for chemical analysis systems, A look into next century's technology or just a fashionable craze?," *Trends in Analytical Chem.*, 10(5):144-149 (1991).
- Manz et al., "Planar chips technology for miniaturization and integration of separation techniques into monitoring systems, Capillary electrophoresis on a chip," *J. Chromatography*, 593:253-258 (1992).
- Manz et al., "Planar Chips Technology for Miniaturization of Separation Systems: A Developing Perspective in Chemical Monitoring," chapter 1, 1-64 (1993).
- Manz et al., "Electroosmotic pumping and electrophoretic separations for miniaturized chemical analysis systems," *J. Micromech. Microeng.*, 4:257-265 (1994).
- Masiakowski et al., "Cloning of cDNA sequences of hormone-regulated genes from the MCF-7 human breast cancer cell line," *Nuc. Acids Res.*, 10(24):7895-7903 (1982).
- Matsumoto et al., "Preliminary Investigation of Micropumping Based on Electrical Control of Interfacial Tension," *IEEE*, pp. 105-110 (1990).
- Matsuzawa et al., "Containment and growth of neuroblastoma cells on chemically patterned substrates," *J. Neurosci. Meth.*, 50:253-260 (1993).
- Matthes et al., "Simultaneous rapid chemical synthesis of over one hundred oligonucleotides on a microscale," *EMBO J.*, 3(4):801-805 (1984).
- McCray et al., "Properties and Uses of Photoreactive Caged Compounds," *Ann. Rev. Biophys. Biophys. Chem.*, 18:239-270 (1989).
- McGall et al., "The Efficiency of Light-Directed Synthesis of DNA Arrays on Glass Substrates," *J. American Chem. Soc.*, 119(22):5081-5090 (1997).
- McGillis, VLSI Technology, Sze, eds., Chapter 7, "Lithography," pp. 267-301 (1983).
- McMurray, J.S., "Solid Phase Synthesis of a Cyclic Peptide Using Fmoc Chemistry," *Tetrahedron Letters*, 32(52):7679-7682 (1991).
- Meinkoth et al., "Review: Hybridization of Nucleic Acids Immobilized on solid Supports," *Analytical Biochem.*, 138:267-284 (1984).
- Melcher et al., "Traveling-Wave Bulk Electroconvection Induced across a Temperature Gradient," *Physics of Fluids*, 10(6):1178-1185 (1967).
- Merrifield, R.B., "Solid Phase peptide Synthesis. I. The Synthesis of a Tetrapeptide," *J. Am. Chem. Soc.*, 85:2149-2154 (1963).
- Michiels et al., "Molecular approaches to genome analysis: a strategy for the construction of ordered overlapping clone libraries," *CABIOS*, 3(3):203-10 (1987).
- Mirzabekov, A.D., "DNA sequencing by hybridization—a megasequencing method and a diagnostic tool?," *TIBTECH*, 12:27-32 (1994).
- Miyada et al., "Oligonucleotide Hybridization Techniques," *Meth. Enzymology*, 154:94-107 (1987).
- Monaco et al., "Human Genome Linking with Cosmids and Yeast Artificial Chromosomes", abstract from CSHS, p. 50, (1989).
- Morita et al., "Direct pattern fabrication on silicone resin by vapor phase electron beam polymerization," *J. Vac. Sci. Technol.*, B1(4):1171-1173 (1983).
- Morrison et al., "Solution-Phase Detection of Polynucleotides Using Interacting Fluorescent Labels and Competitive Hybridization," *Anal. Biochem.*, 183:231-244 (1989).

US 6,403,320 B1

Page 9

- Munegumi et al., "thermal Synthesis of Polypeptides from N-Boc-Amino Acid (Aspartic Acid, β -Aminoglutaric Acid) Anhydrides," *Chem. Letters*, pp. 1643-1646 (1988).
- Mutter et al., "Impact of Conformation on the Synthetic Strategies for Peptide Sequences," pp. 217-228 from Chemistry of Peptides and Proteins, vol. 1, Proceedings of the Third USSR-FRG Symp., in USSR (1982).
- Nakamori et al., "A Simple and Useful Method for Simultaneous Screening of Elevated Levels of Expression of a Variety of Oncogenes in Malignant Cells," *Jpn. J. Cancer Res.*, 79:1311-1317 (1988).
- Nederlof et al., "Multiple Fluorescence In Situ Hybridization," *Cytometry*, 11:126-131 (1990).
- Nederlof et al., "Three-Color Fluorescence In Situ Hybridization for the Simultaneous Detection of Multiple Nucleic Acid Sequences," *Cytometry*, 10:20-27 (1989).
- Nizetic et al., "An improved bacterial colony lysis procedure enables direct DNA hybridisation using short (10, 11 bases) oligonucleotides to cosmid," *Nuc. Acids Res.*, 19(1):182 (1990).
- Nizetic et al., "Construction, arraying, and high-density screening of large insert libraries of human chromosomes X and 21: their potential use as reference libraries," *PNAS*, 88:3233-3237 (1991).
- Nyborg, W., "Acoustic Streaming," chapter 11 pp. 265-329 from Physical Acoustics, Principles and Methods, Mason, eds., vol. II, part B, Academic Press, New York and London (1965).
- Ocvirk et al., "High Performance Liquid Chromatography Partially Integrated onto a Silicon Chip," *Analyt. Meth. Instrumentation*, 2(2):74-82 (1995).
- Ohtsuka et al., "Studies on transfer ribonucleic acids and related compounds. IX Ribonucleic oligonucleotide synthesis using a photosensitive 0-nitrobenzyl protection at the 2'-hydroxyl group," *Nuc. Acids Res.*, 1(10):1351-1357 (1974).
- Olefirowicz et al., "Capillary Electrophoresis for Sampling Single Nerve Cells," *Chimia*, 45(4):106-108 (1991).
- Olson et al., "Random-clone strategy for genomic restriction mapping in yeast," *PNAS*, 83:7826-7830 (1986).
- Patchornik et al., "Photosensitive Protecting Groups," *J. Am. Chem. Soc.*, 92(21):6333-6335 (1970).
- Patent Abstracts of Japan from EPO, Abst. 13:557, JP 1-233 447 (1989).
- Pease et al., "Light-generated oligonucleotide arrays for rapid DNA sequence analysis," *PNAS*, 91:5022-26 (1994).
- Pevzner, P.A., "DNA Physical Mapping and Alternating Eulerian Cycles in Colored Grapes," *Algorithmica*, 13(1-2):77-105 (1995).
- Pevzner, et al., "Multiple Filtration and Approximate Pattern Matching," *Algorithmica*, 13(1-2):135-154 (1995).
- Pevzner et al., "Generalized Sequence Alignment and Duality," *Adv. Applied Math.*, 14:139-171 (1993).
- Pevzner, P.A., "1-Tuple DNA Sequencing: Computer Analysis," *J. Biomol. Struct. Dynam.*, 7(1):63-69 (1989).
- Pfahler et al., "Liquid Transport in Micron and Submicron Channels," *Sensors and Actuators*, A21-A23:431-4 (90).
- Pfeifer et al., "Genomic Sequencing and Methylation Analysis by Ligation Mediated PCR," *Science*, 246:810-813 (1989).
- Pidgeon et al., "Immobilized Artificial Membrane Chromatography: Supports Composed of Membrane Lipids," *Anal. Biochem.*, 176:36-47 (89).
- Pillai, V.N., "Photoremovable Protecting Groups in Organic Synthesis," *Synthesis*, pp. 1-26 (1980).
- Pillai et al., "3-Nitro-4-Aminomethylbenzoyl derivative von Polyethylenglykolen: Eine neue Klasse von Photosensitiven loslichen Polymeren Tragern zur Synthese von C-terminalen Peptidamiden," *Tetrah. Ltr.*, #36 p. 3409-3412 (1979).
- Pillai et al., "Synthetic Hydrophilic Polymers, Biomedical and Chemical Applications," *Naturwissenschaften*, 68:558-566 (1981).
- Pirrung et al., "Proofing of Photolithographic DNA Synthesis with 3'.5'-Dimethoxybenzoinyloxycarbonyl-Protected Deoxynucleoside Phosphoramidites," *J. Org. Chem.*, 63(2):241-246 (1998).
- Pirrung et al., "Comparison of Methods for Photochemical Phosphoramidite-Based DNA Synthesis," *J. Org. Chem.*, 60:6270-6276 (1995).
- Ploax et al., "Cyclization of peptides on a solid support," *Int. J. Peptide Protein Research*, 29:162-169 (1987).
- Polsky-Cynkin et al., "Use of DNA Immobilized on Plastic and Agarose Supports to Detect DNA by Sandwich Hybridization," *Clin. Chem.*, 31(9):1428-1443 (1985).
- Poustka et al., "Molecular Approaches to Mammalian Genetics," Cold Spring Harbor Symposia on Quantitative Biology, 51:131-139 (1986).
- Purushothaman et al., "Synthesis of 4,5-diarylimidazole-2 thiones and their photoconversion to bis(4, 5-diarylimidazol-2-yl) sulphides," *Ind. J. Chem.*, 29B:18-21 (1990).
- Quesada et al., "High-Sensitivity DNA Detection with a Laser-Excited Confocal Fluorescence Gel Scanner," *Biotechniques*, 10:616 (1991).
- Reichmanis et al., *J. Polymer Sci. Polymer Chem. Edition*, 23:1-8 (1985).
- Renz et al., "A colorimetric method for DNA hybridization," *Nuc. Acids Res.*, 12(8):3435-3445 (1984).
- Richter et al., "An Electrohydrodynamic Micropump," *IEEE*, pp. 99-104 (1990).
- Richter et al., "Electrodynamic Pumping and Flow Measurement," *IEEE*, pp. 271-276 (1991).
- Richter et al., "A Micromachined electrohydrodynamic (EHD) pump," *Sensors and Actuators*, A29:159-168 (91).
- Robertson et al., "A General and Efficient Route for Chemical Aminoacylation of Transfer RNAs," *J. Am. Chem. Soc.*, 113:2722-2729 (1991).
- Rodda et al., "The Antibody Response to Myoglobin-I. Systematic Synthesis of Myoglobin Peptides Reveals Location and Substructure of Species-Dependent Continuous Antigenic Determinants," *Mol. Immunol.*, 23(6):603-610 (1986).
- Rodgers, R.P., "Data Processing of Immunoassay Results," Manual of Clin. Lab. Immunol., 3rd ed., ch. 15, pp. 82-87 (1986).

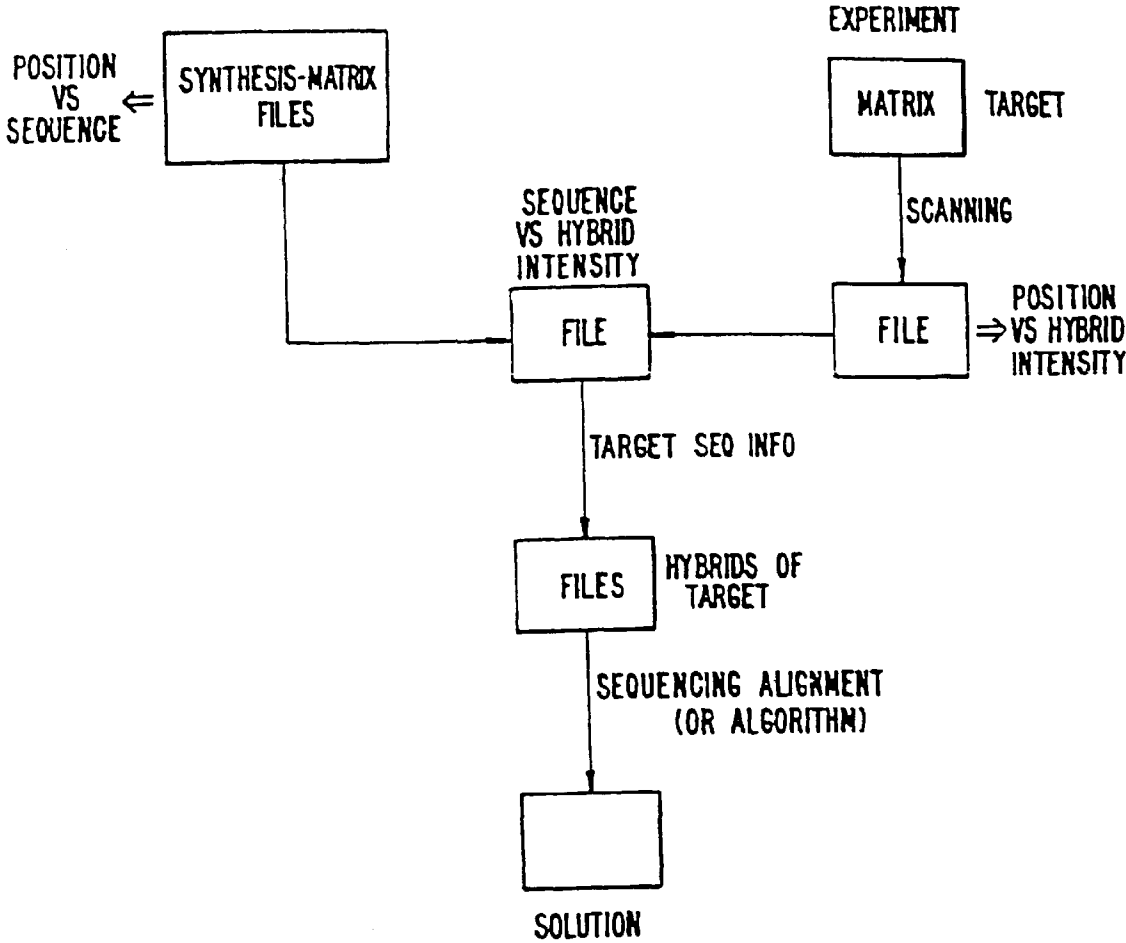


FIGURE 1

U.S. Patent

Jun. 11, 2002

Sheet 2 of 27

US 6,403,320 B1

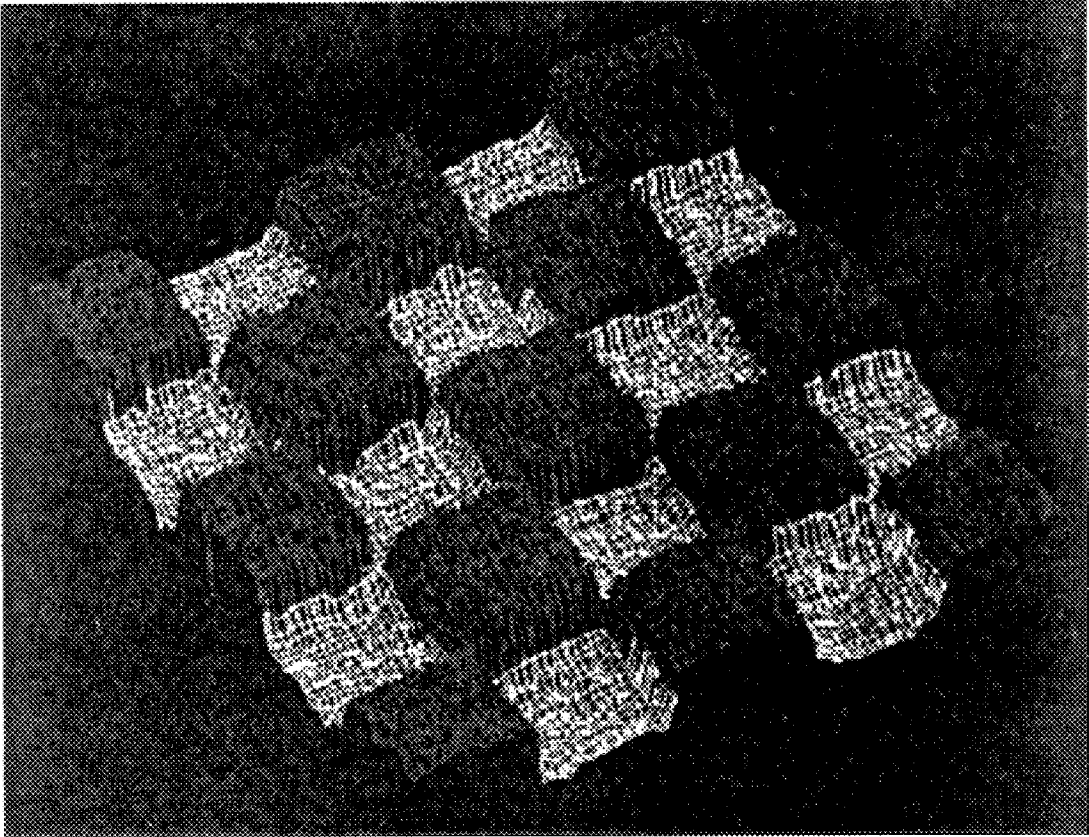


Figure 2

U.S. Patent

Jun. 11, 2002

Sheet 3 of 27

US 6,403,320 B1

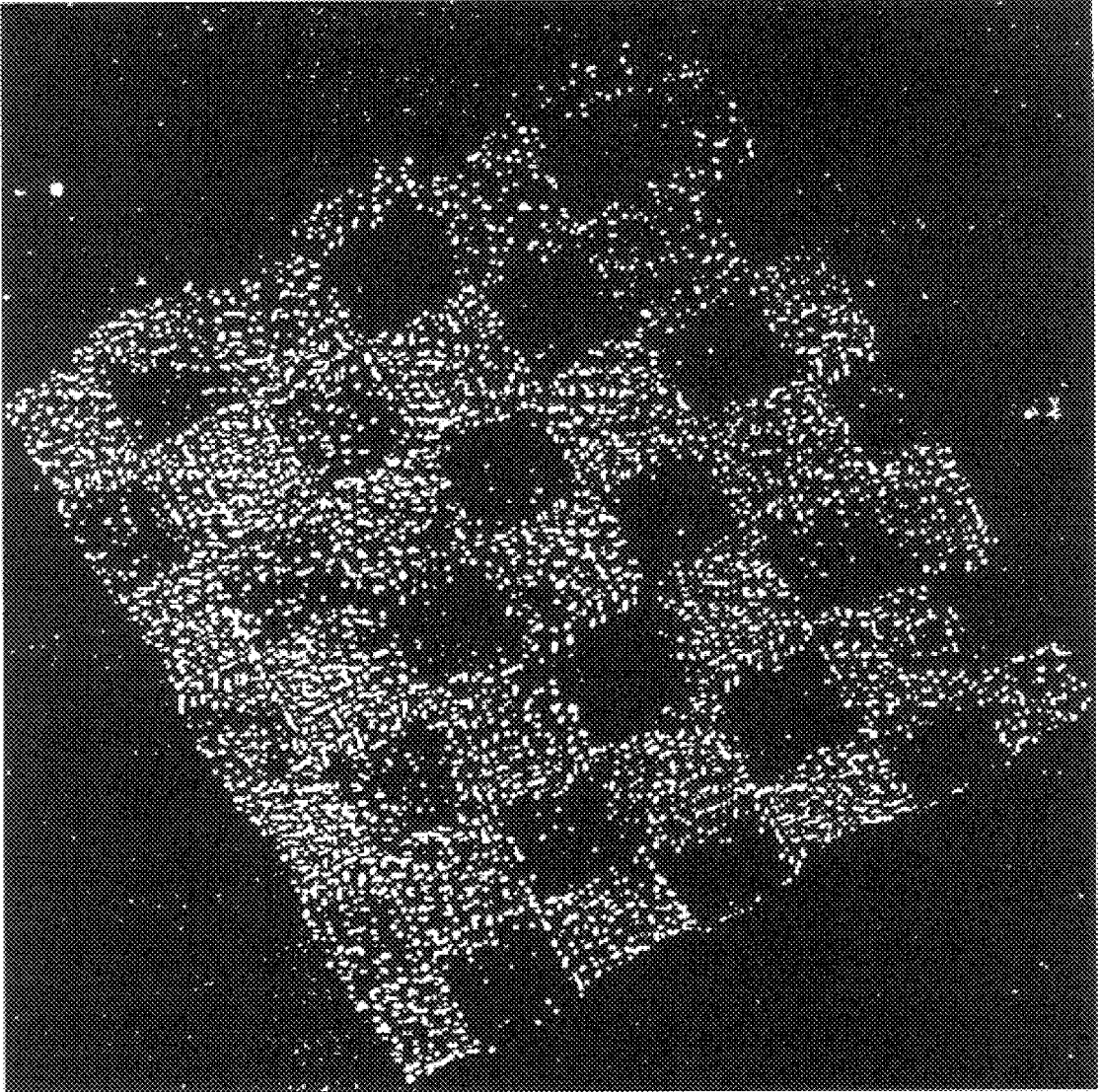


figure 3

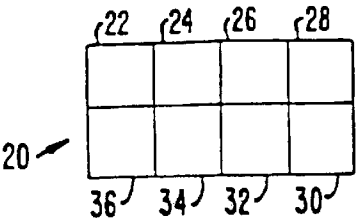


FIG. 4A

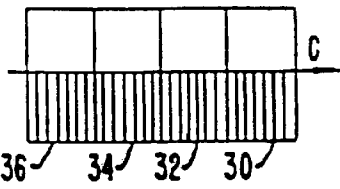


FIG. 4B

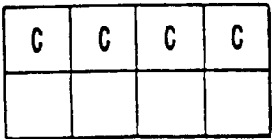


FIG. 4C

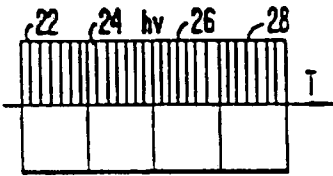


FIG. 4D

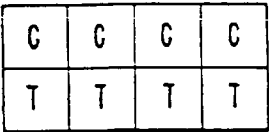


FIG. 4E

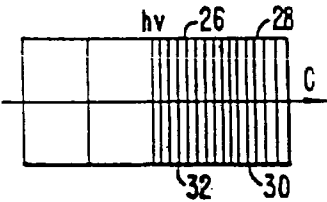


FIG. 4F

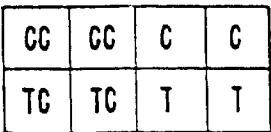


FIG. 4G

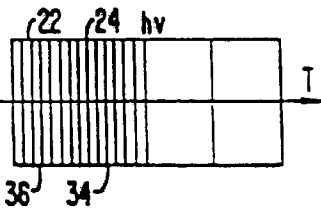


FIG. 4H

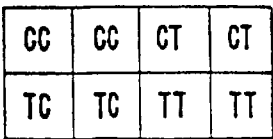


FIG. 4I

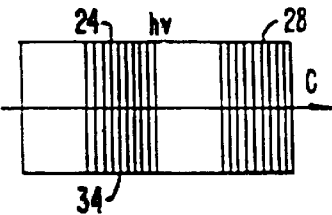


FIG. 4J

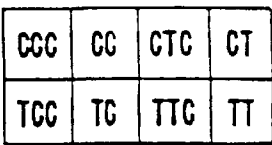


FIG. 4K

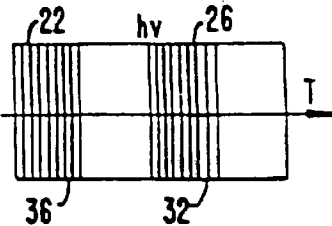


FIG. 4L

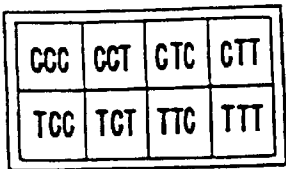


FIG. 4M

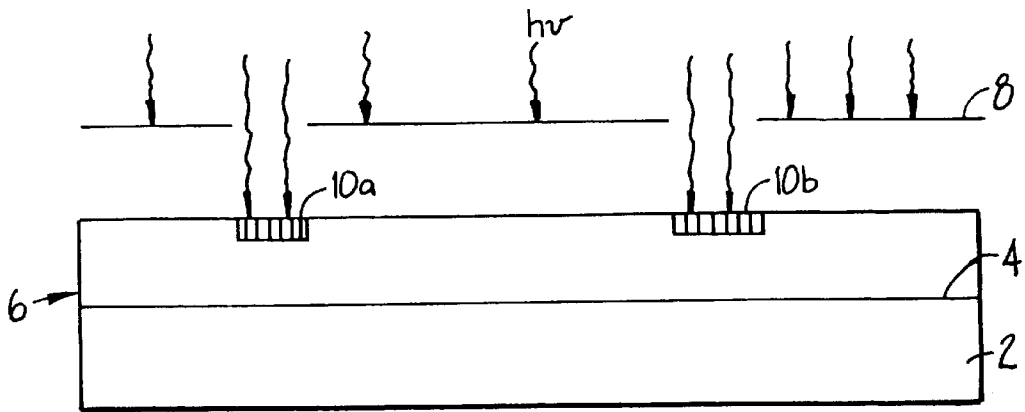


FIGURE 5

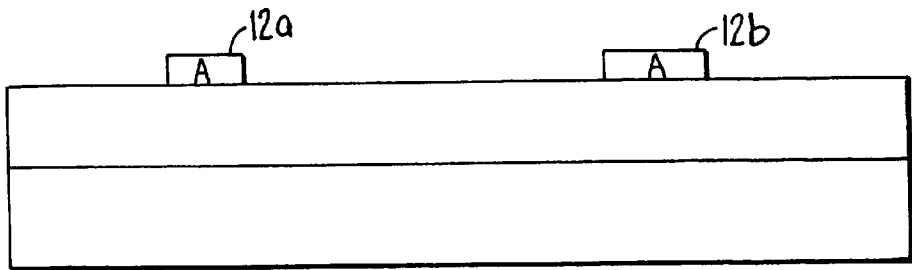


FIGURE 6

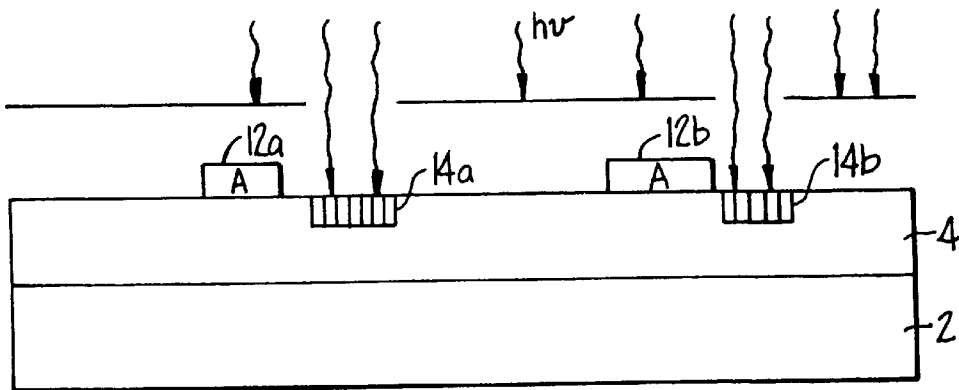


FIGURE 7

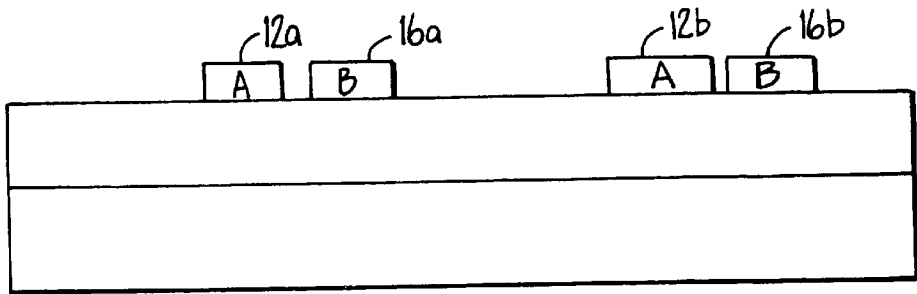
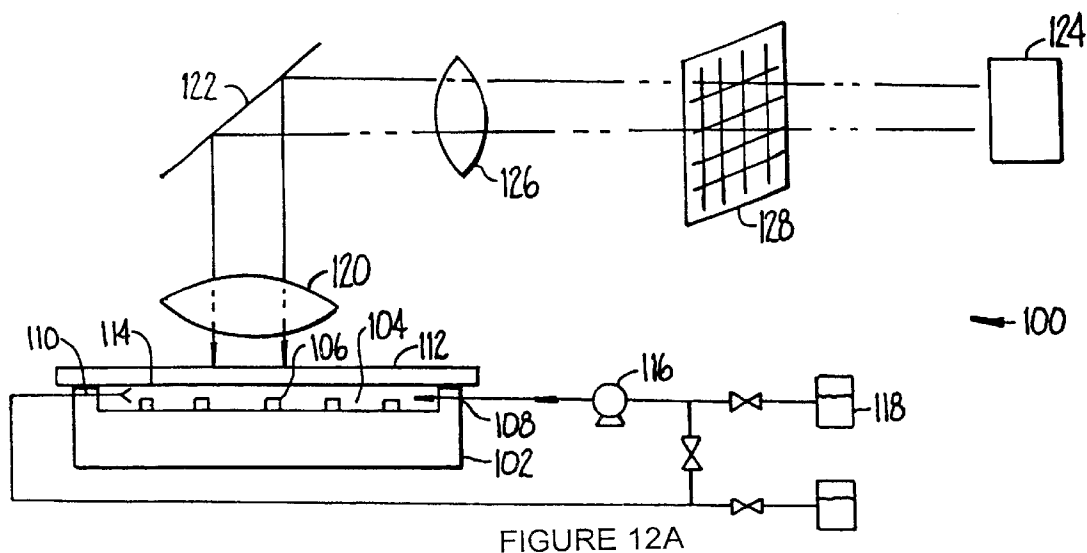
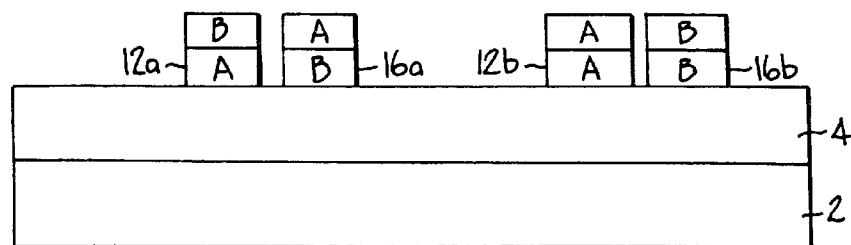
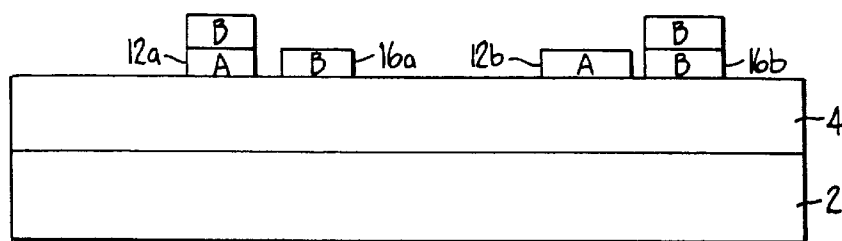
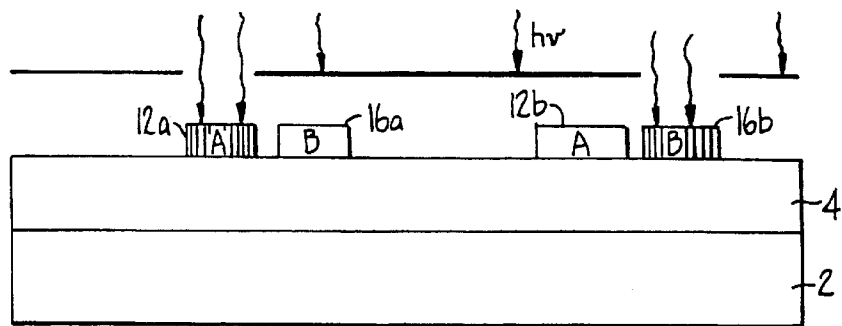


FIGURE 8



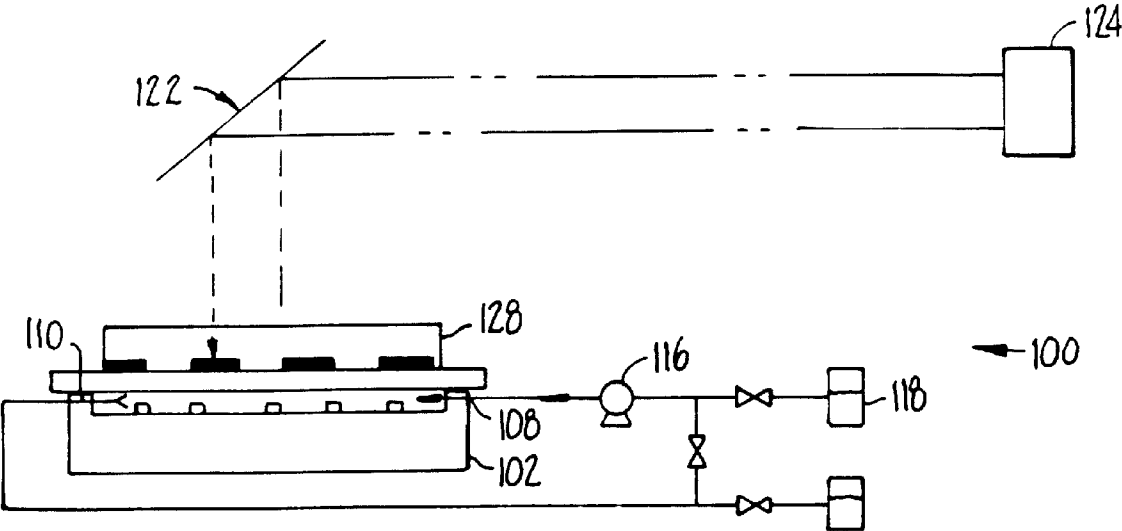


FIG. 12B

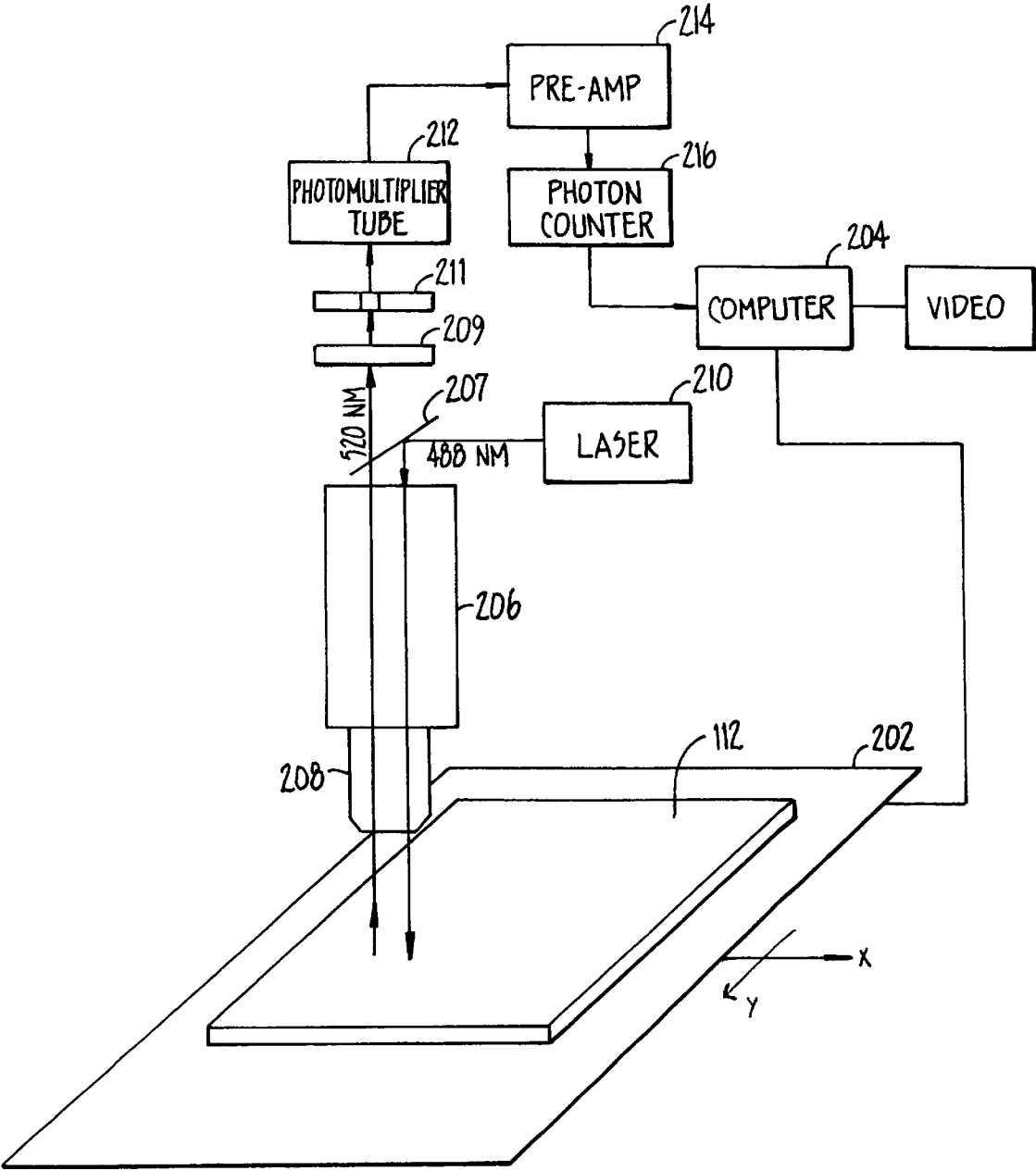


FIGURE 13

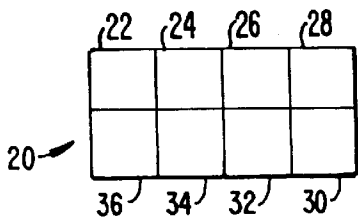


FIG. 14A

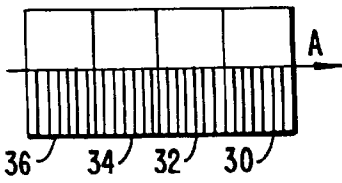


FIG. 14B

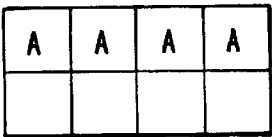


FIG. 14C

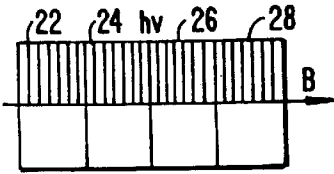


FIG. 14D

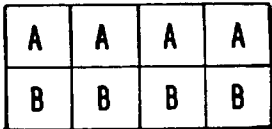


FIG. 14E

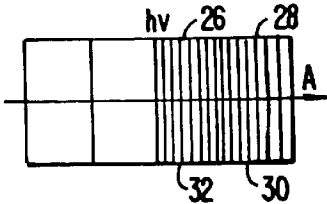


FIG. 14F

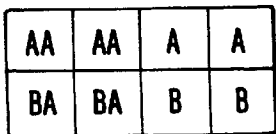


FIG. 14G

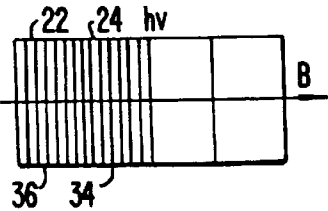


FIG. 14H

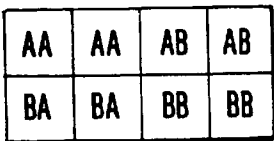


FIG. 14I

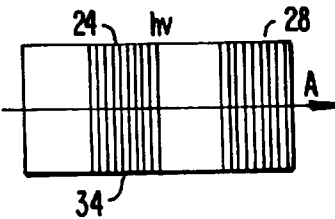


FIG. 14J

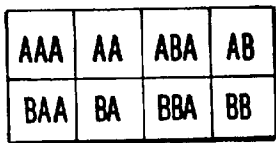


FIG. 14K

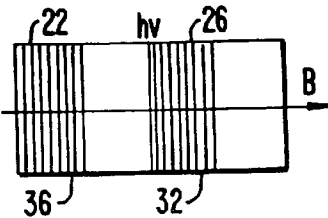


FIG. 14L

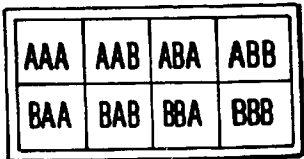


FIG. 14M

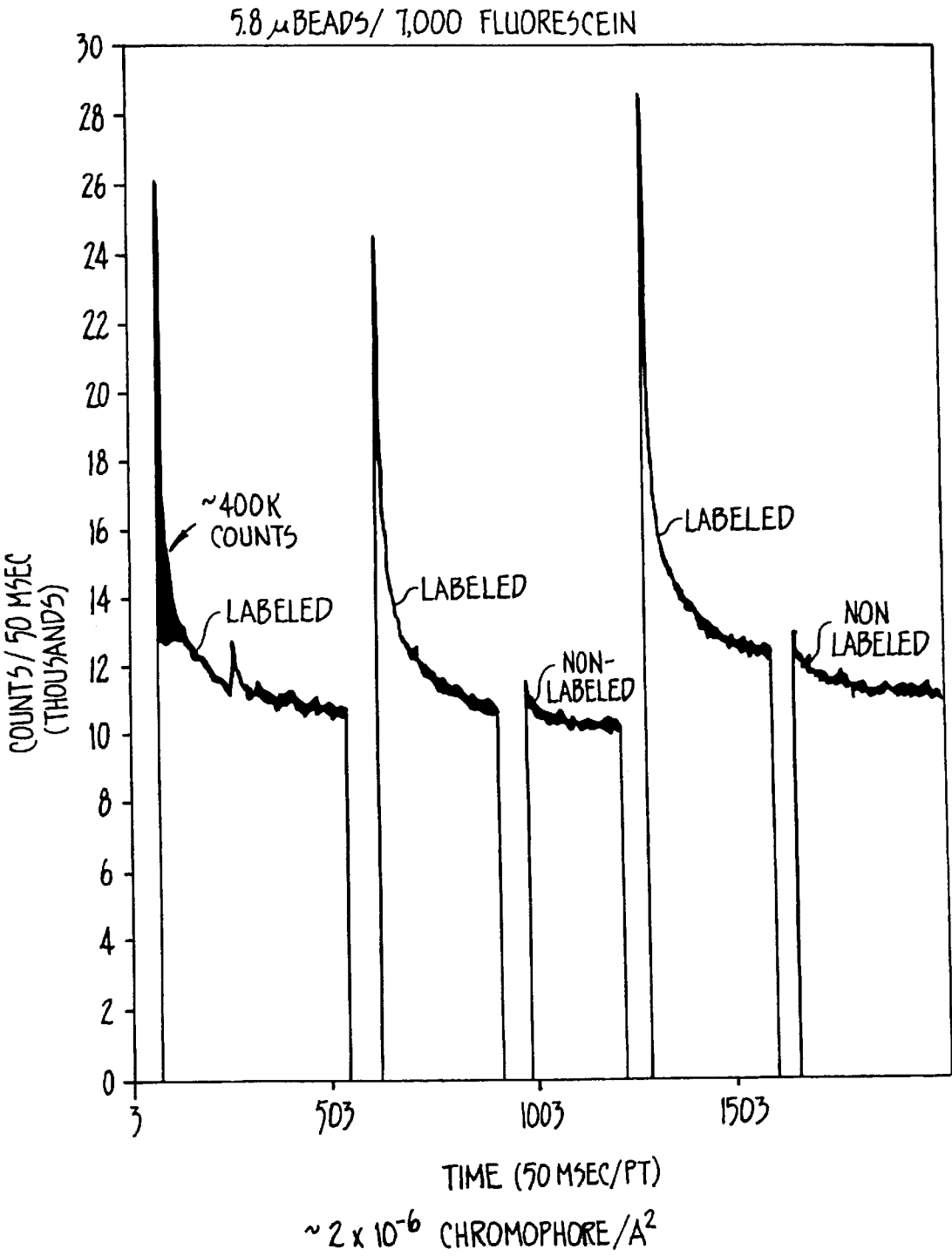


FIGURE 15A

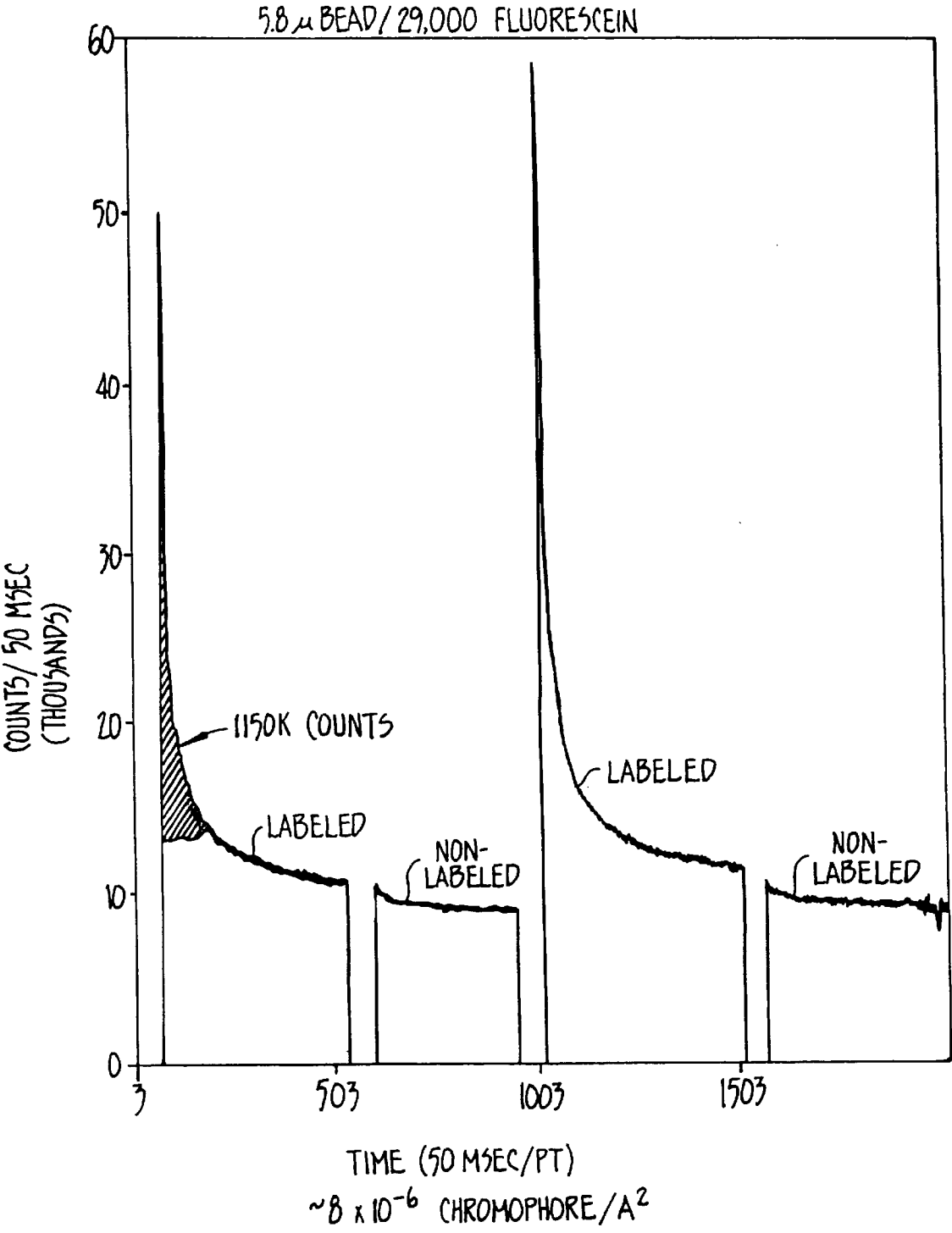


FIGURE 15B

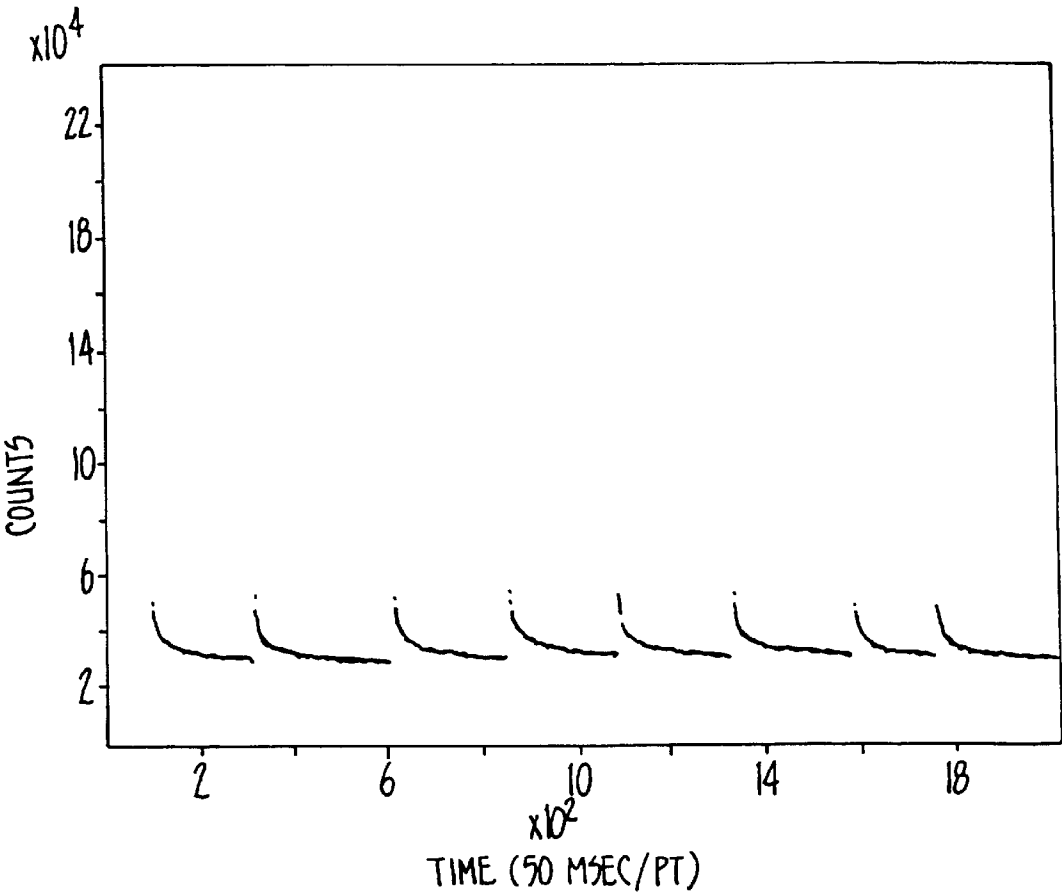


FIGURE 16A

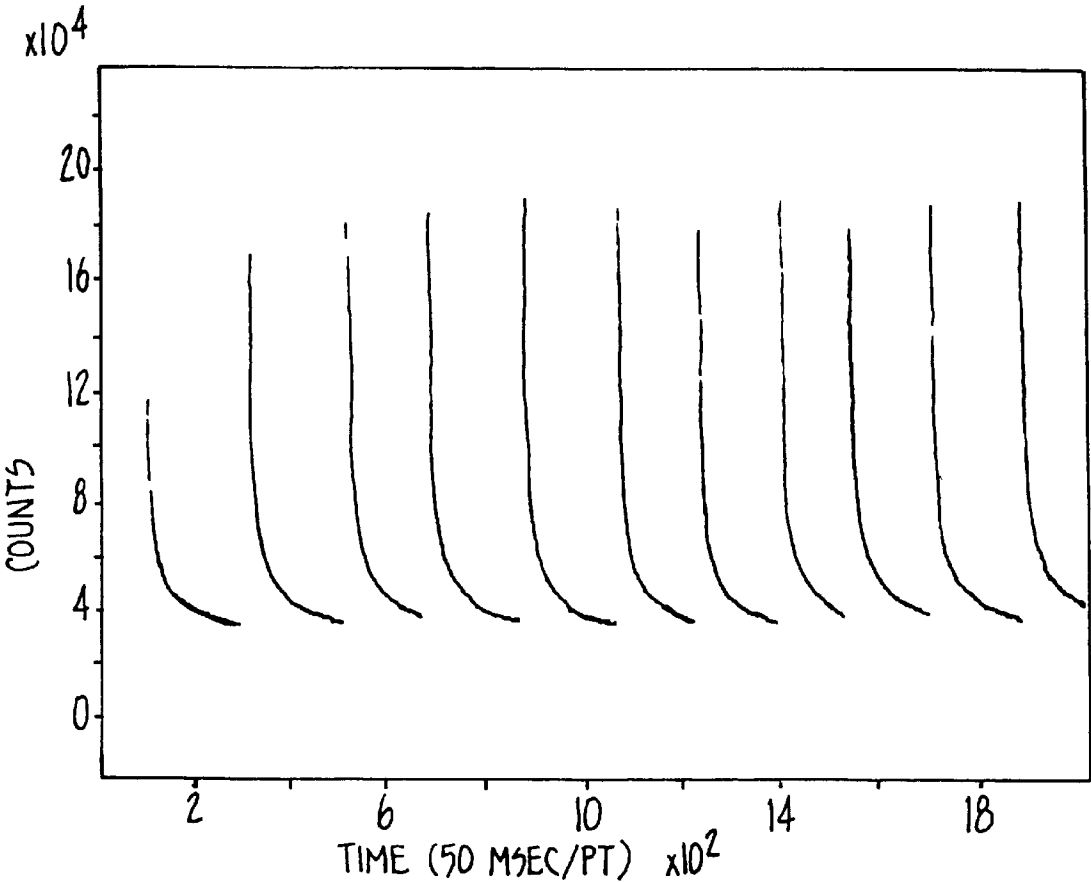


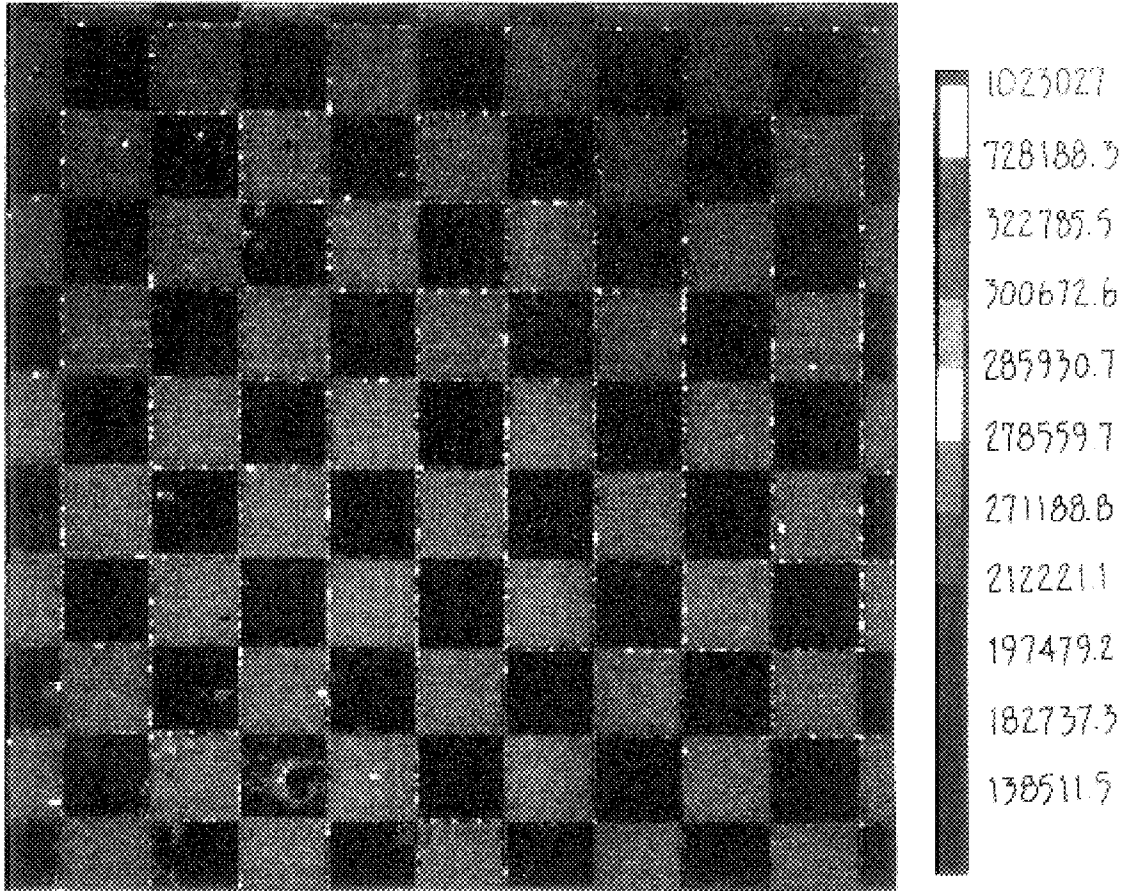
FIGURE 16B

U.S. Patent

Jun. 11, 2002

Sheet 14 of 27

US 6,403,320 B1



MEAN: 285930.7
VAR: 2.173242E+10
 σ : 147419.2

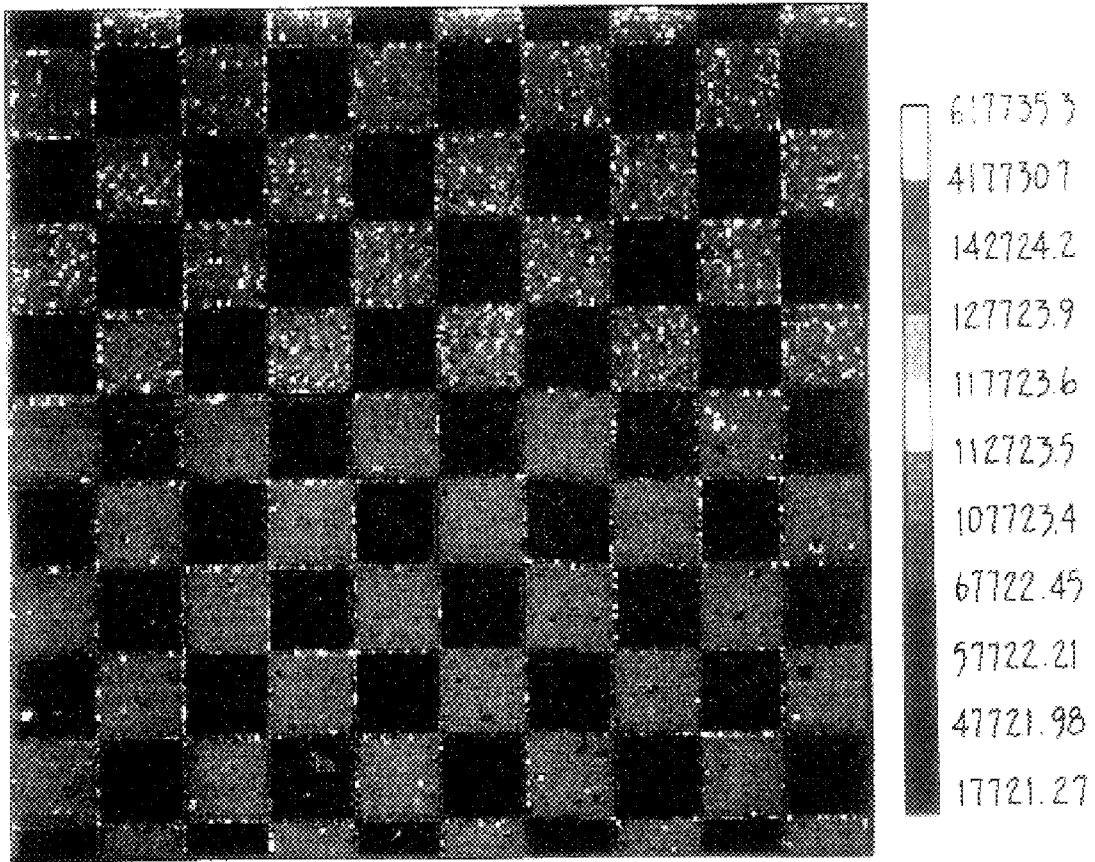
FIG. 17A

U.S. Patent

Jun. 11, 2002

Sheet 15 of 27

US 6,403,320 B1



MEAN: 117723.6
VAR: 1.000047E+10
 σ : 100002.3

FIG. 17B

U.S. Patent

Jun. 11, 2002

Sheet 16 of 27

US 6,403,320 B1

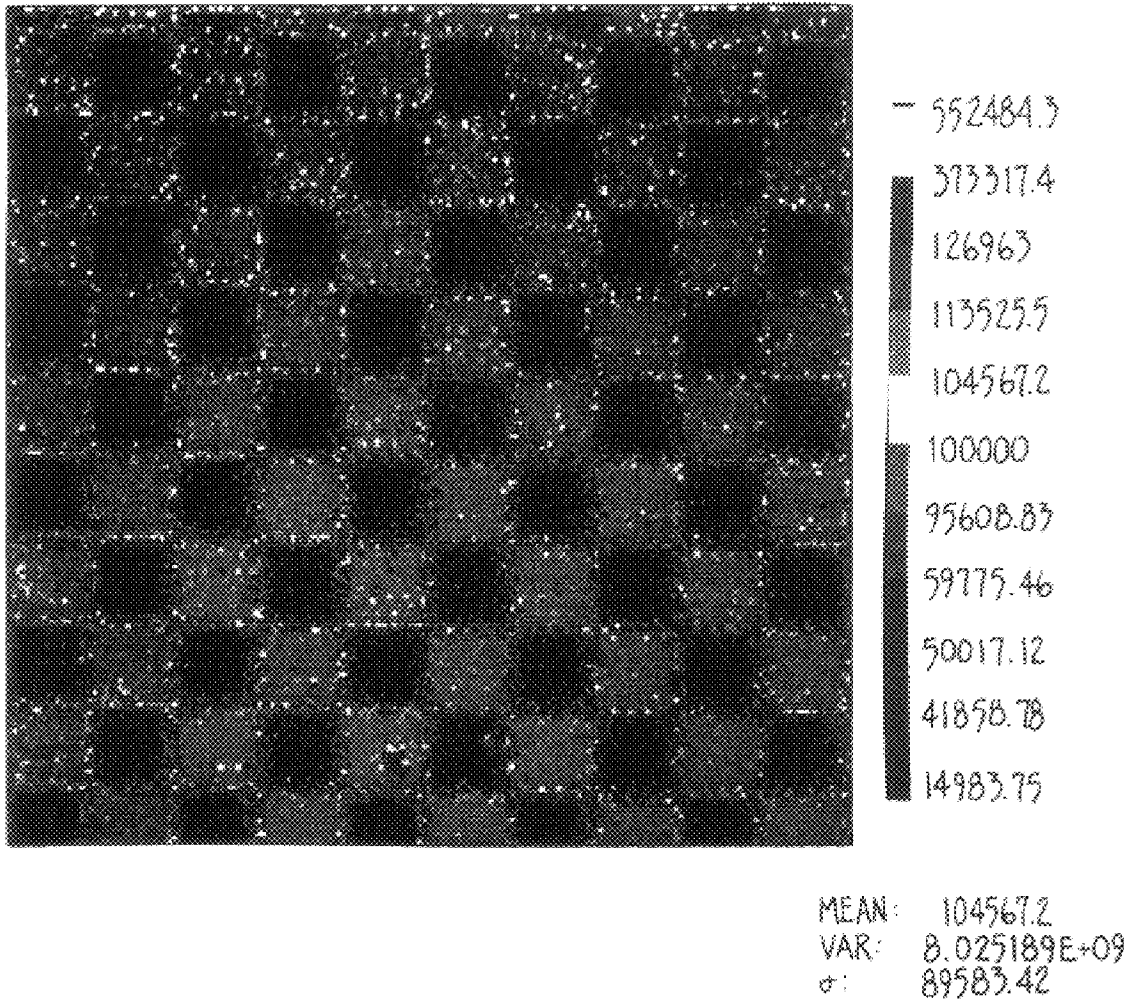


FIG. 17C

U.S. Patent

Jun. 11, 2002

Sheet 17 of 27

US 6,403,320 B1

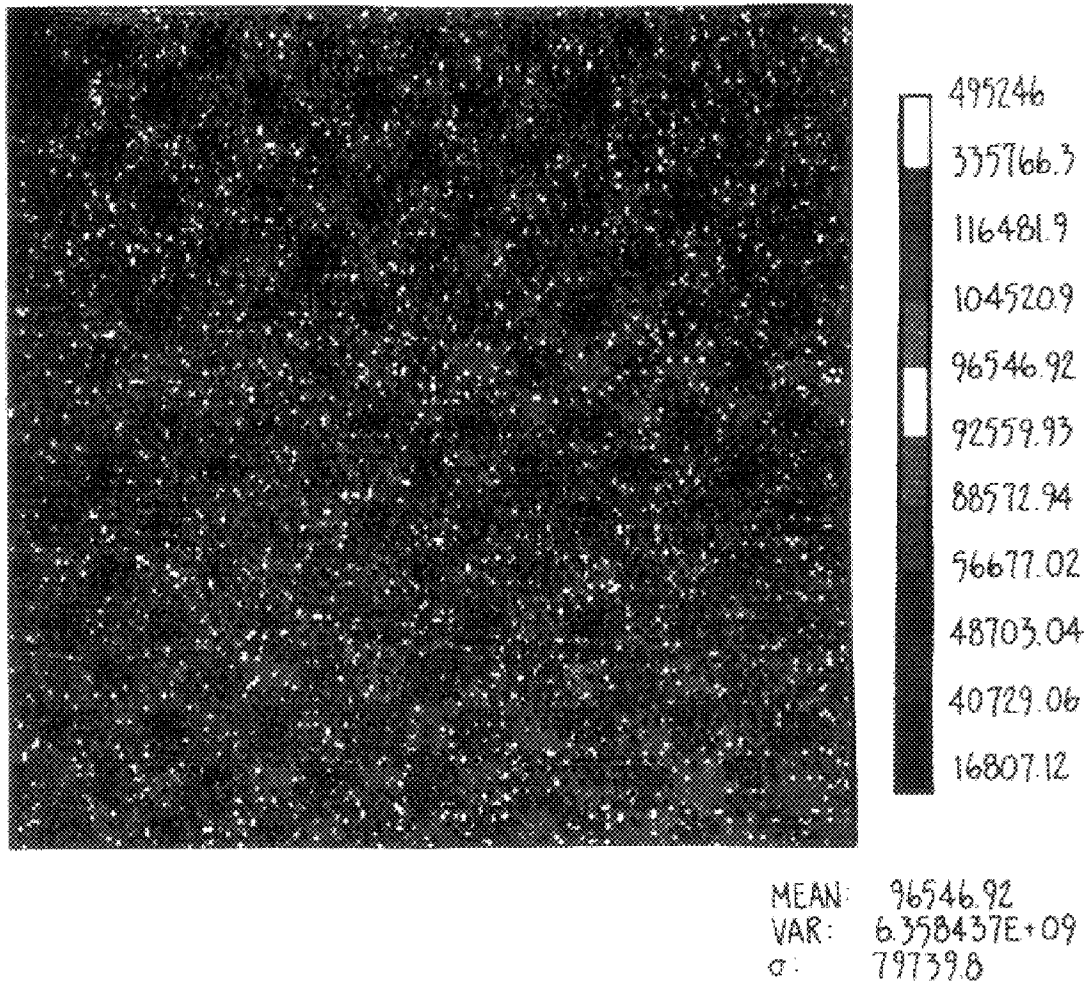


FIG. 17D

U.S. Patent

Jun. 11, 2002

Sheet 18 of 27

US 6,403,320 B1

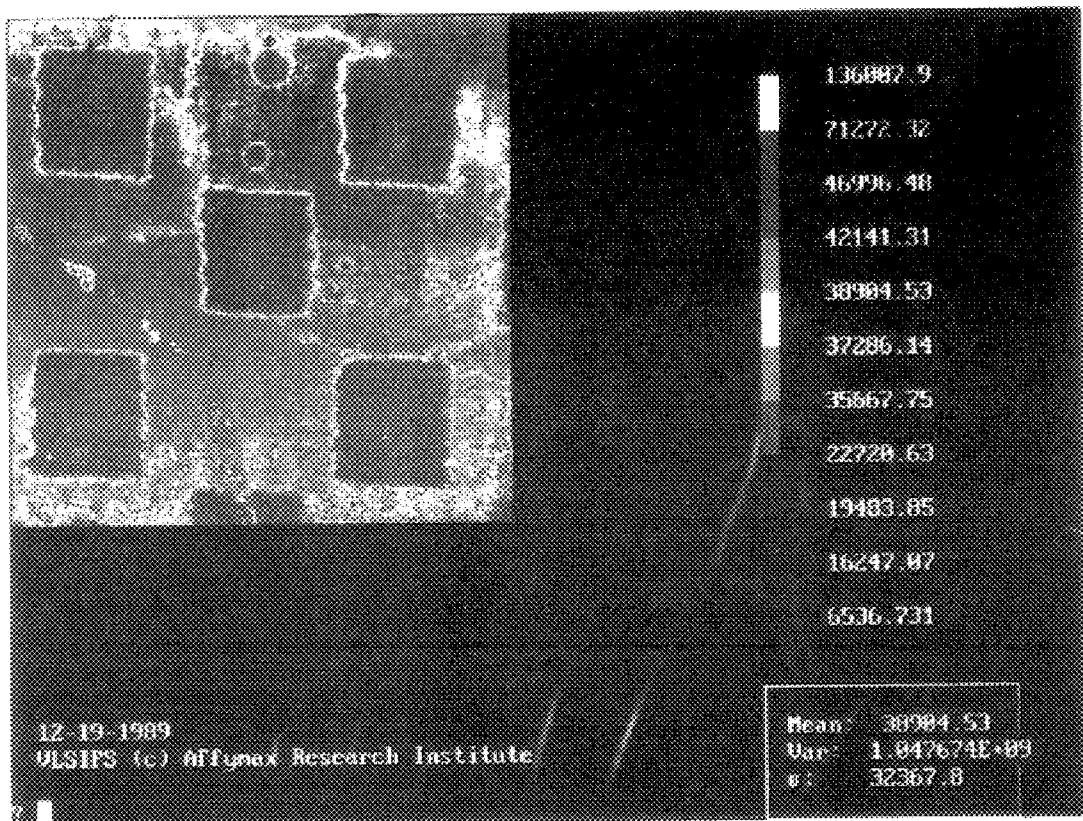


FIG. 18

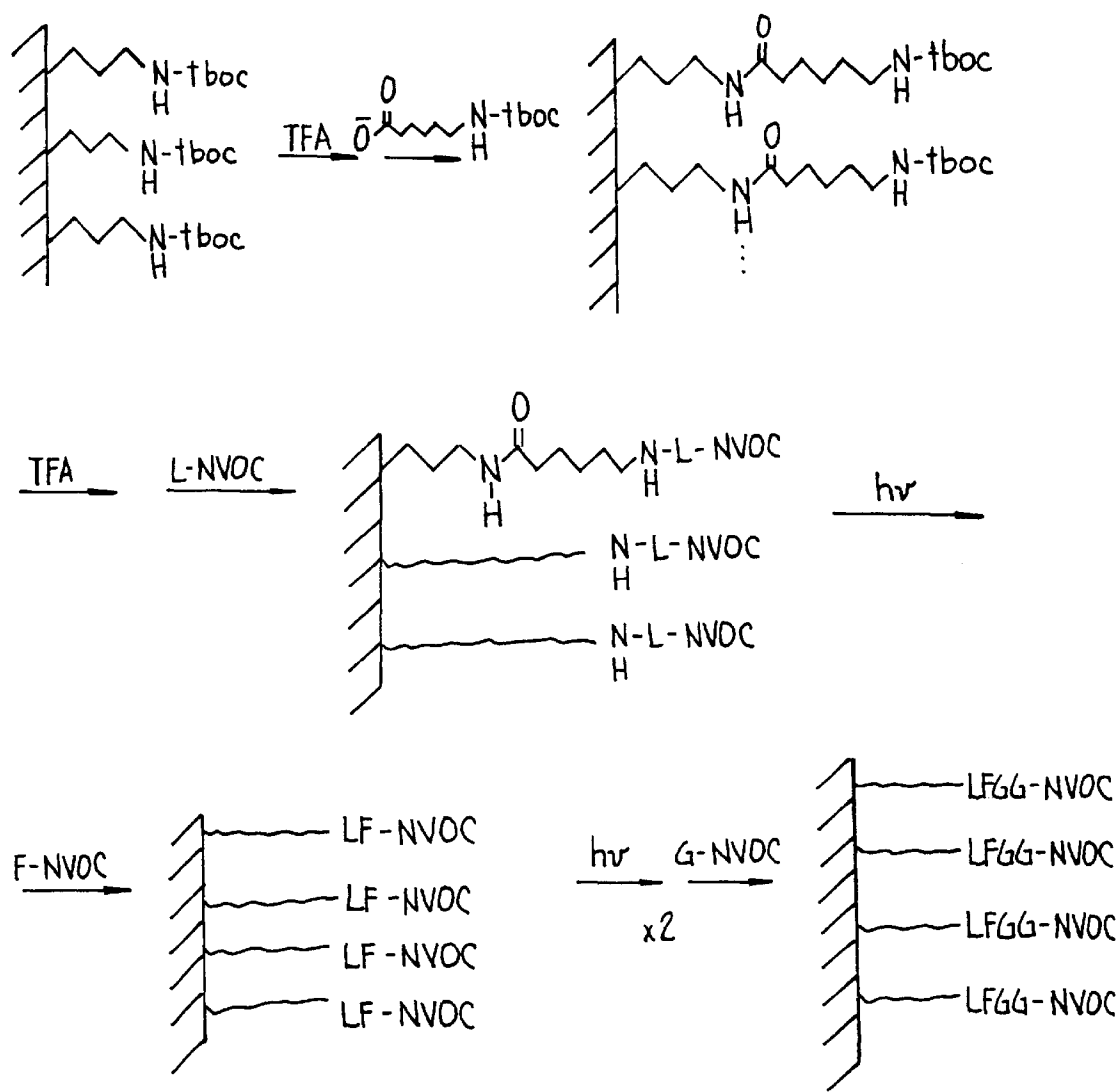


FIGURE 19A

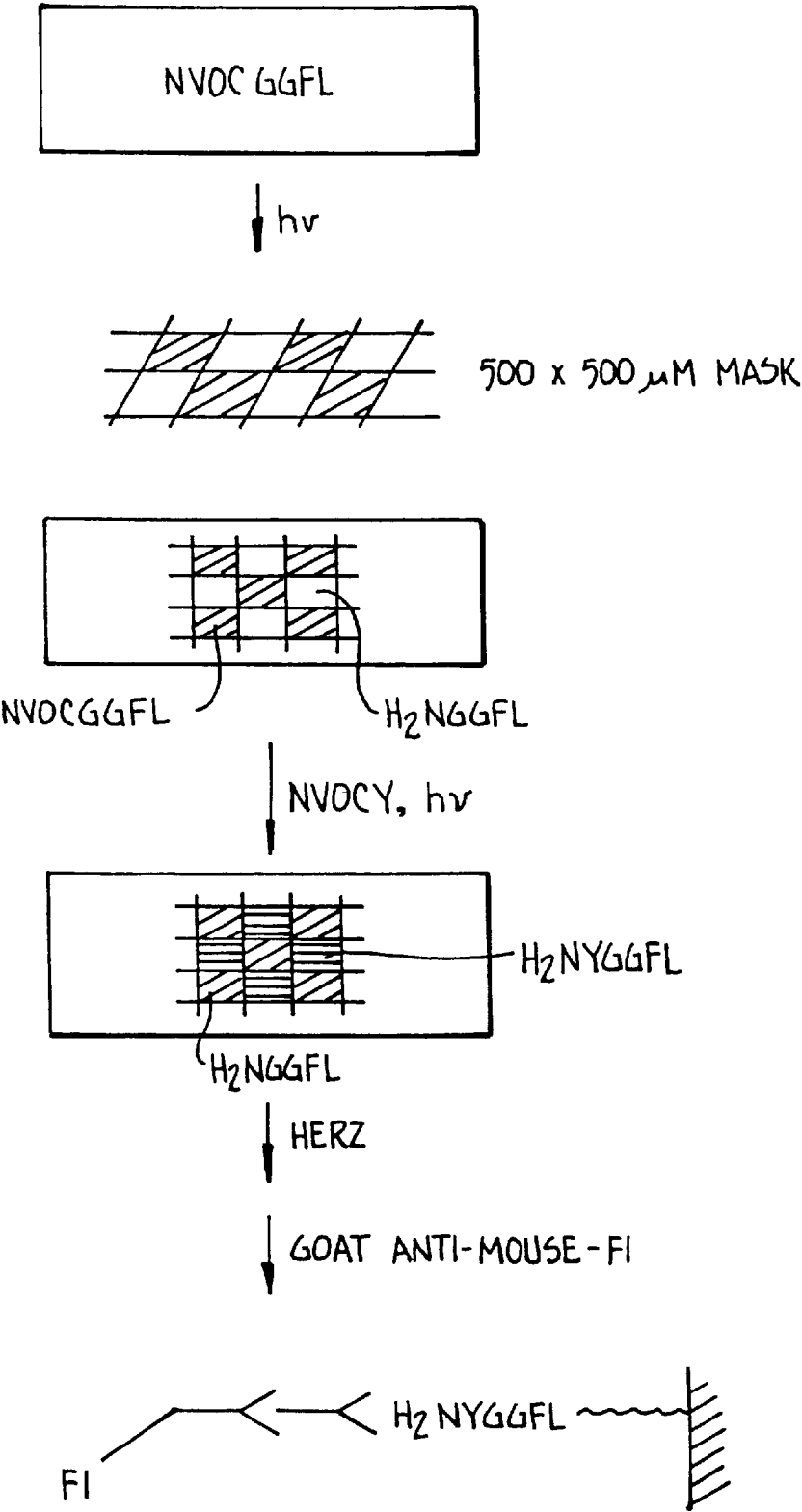


FIGURE 19B

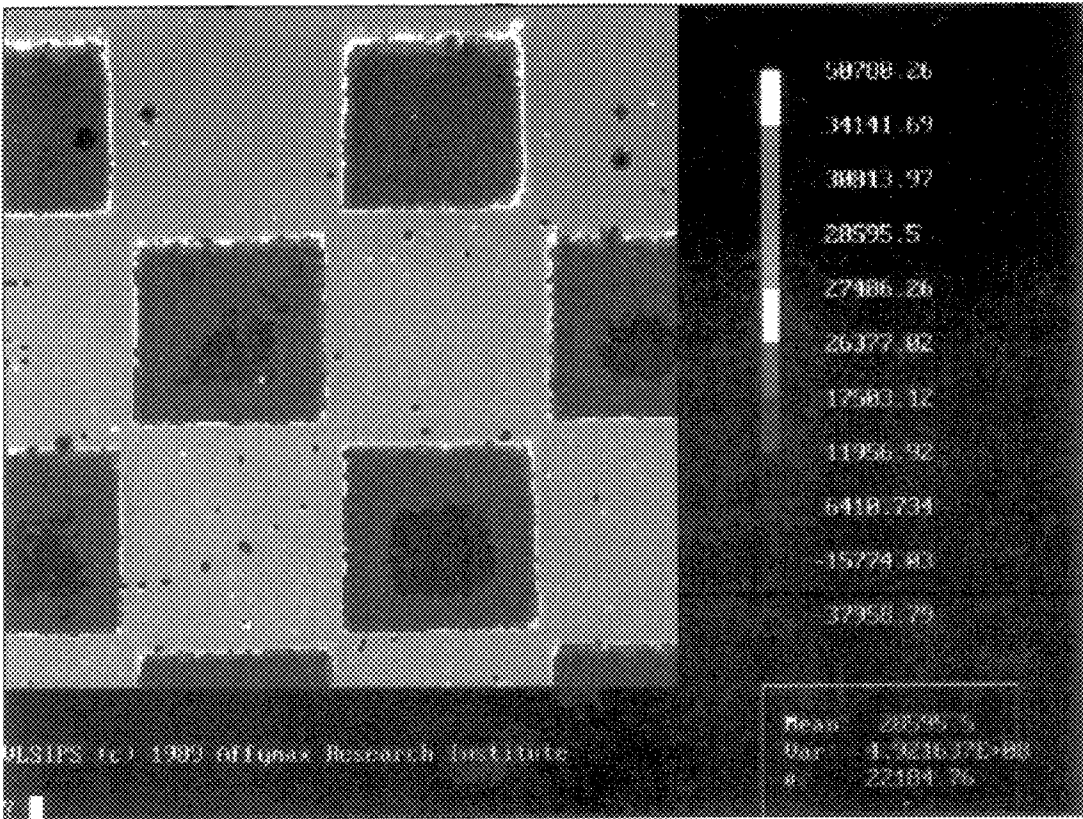


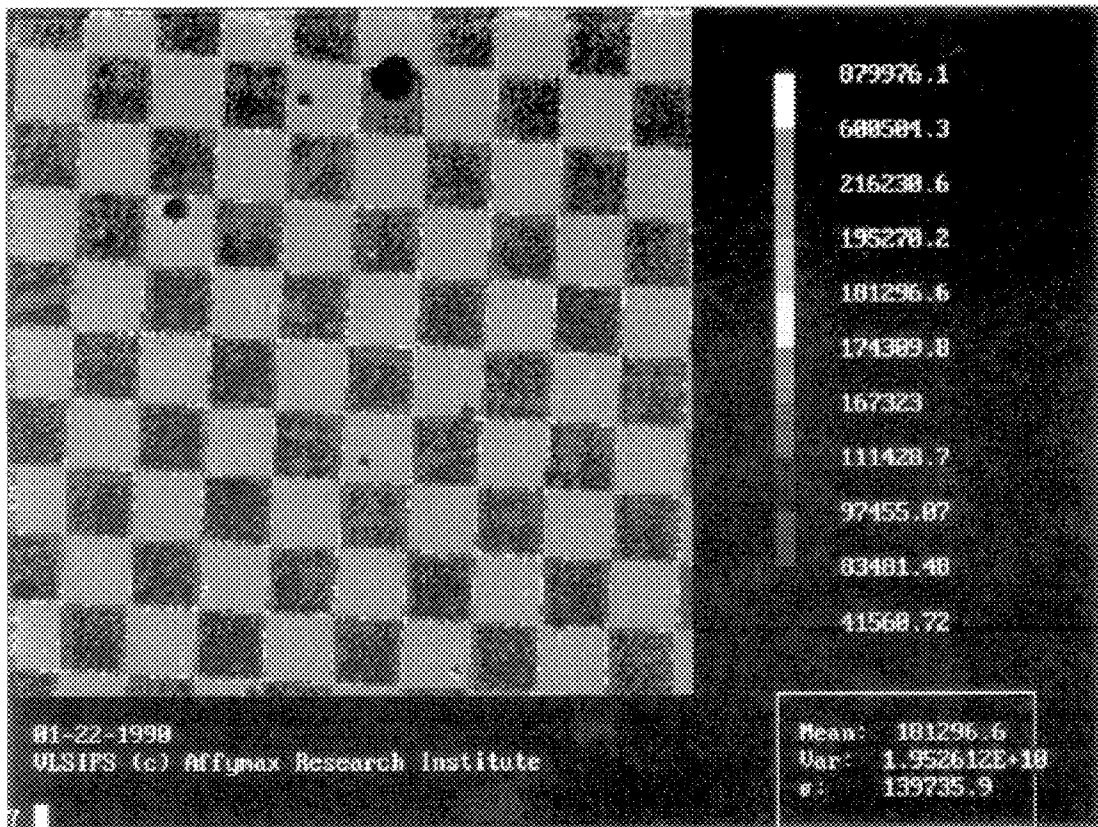
FIG. 19C

U.S. Patent

Jun. 11, 2002

Sheet 22 of 27

US 6,403,320 B1



U.S. Patent

Jun. 11, 2002

Sheet 23 of 27

US 6,403,320 B1

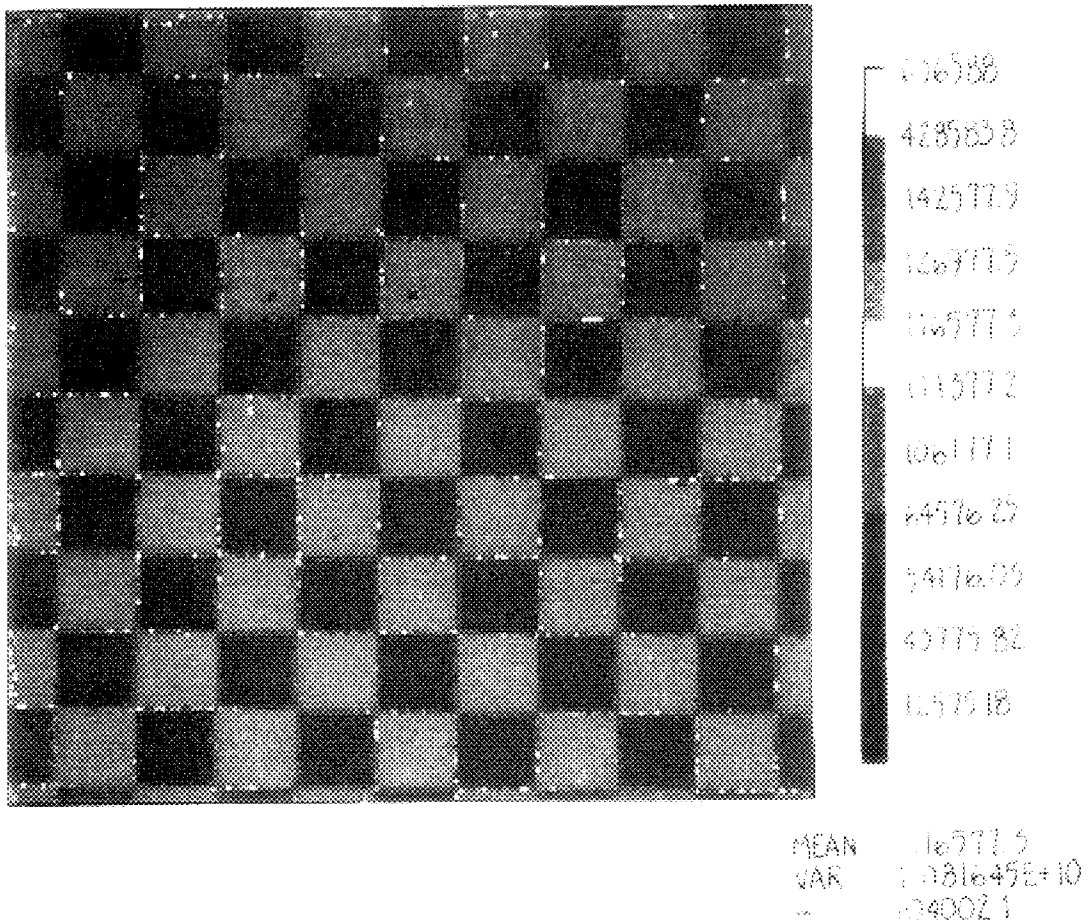


FIG. 20

U.S. Patent

Jun. 11, 2002

Sheet 24 of 27

US 6,403,320 B1

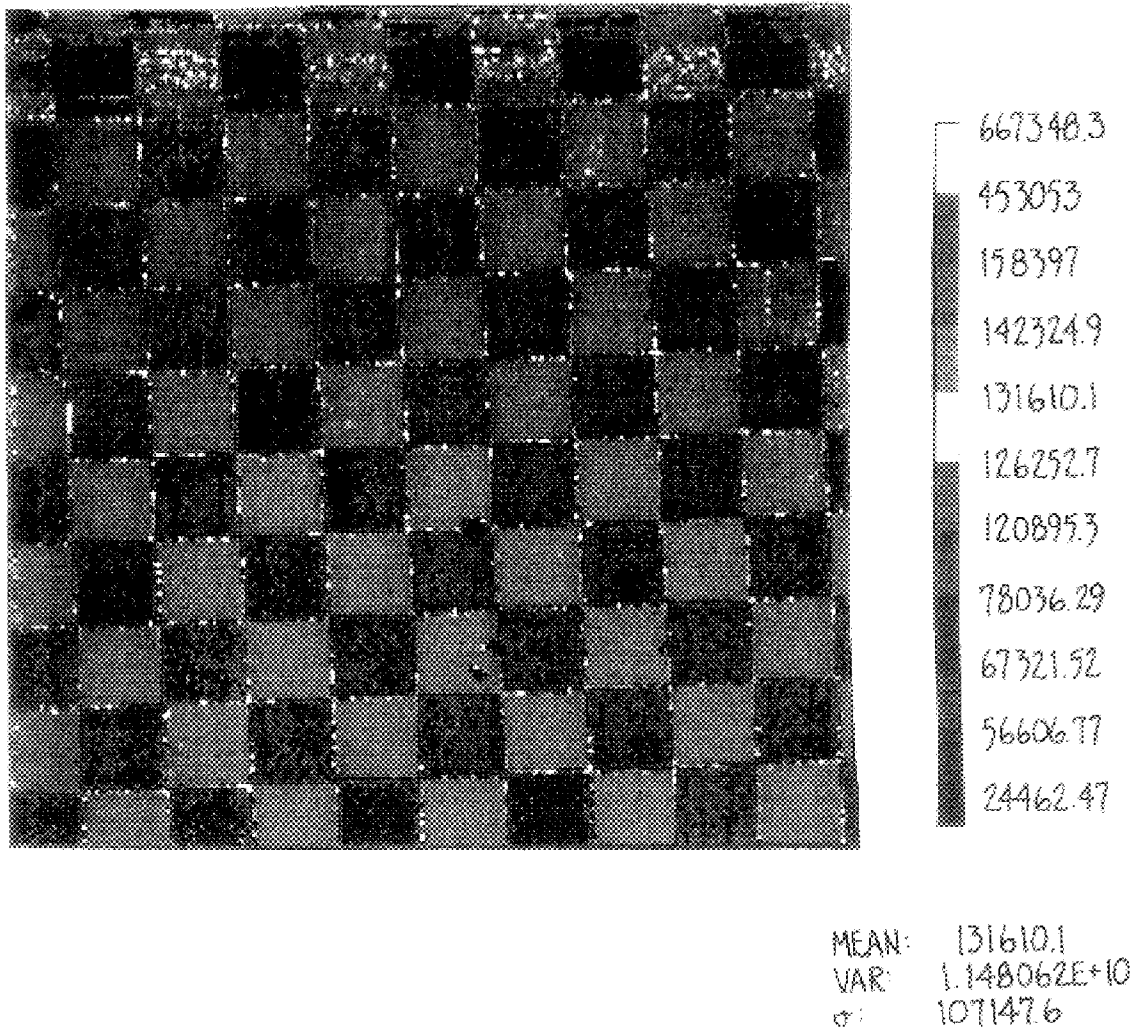


FIG. 21

P	A	S	G	
<u>L</u> P GFL	<u>L</u> A GFL	<u>L</u> S GFL	<u>L</u> G GFL	L
<u>F</u> P GFL	<u>F</u> A GFL	<u>F</u> S GFL	<u>F</u> G GFL	F
<u>W</u> P GFL	<u>W</u> A GFL	<u>W</u> S GFL	<u>W</u> G GFL	W
<u>Y</u> P GFL	<u>Y</u> A GFL	<u>Y</u> S GFL	<u>Y</u> G GFL	Y

L SET

FIGURE 22A

p	a	s	G	
<u>Y</u> p GFL	<u>Y</u> a GFL	<u>Y</u> s GFL	<u>Y</u> G GFL	Y
<u>f</u> p GFL	<u>f</u> a GFL	<u>f</u> s GFL	<u>f</u> G GFL	f
<u>w</u> p GFL	<u>w</u> a GFL	<u>w</u> s GFL	<u>w</u> G GFL	w
<u>y</u> p GFL	<u>y</u> a GFL	<u>y</u> s GFL	<u>y</u> G GFL	y

D SET

FIGURE 22B

U.S. Patent

Jun. 11, 2002

Sheet 26 of 27

US 6,403,320 B1

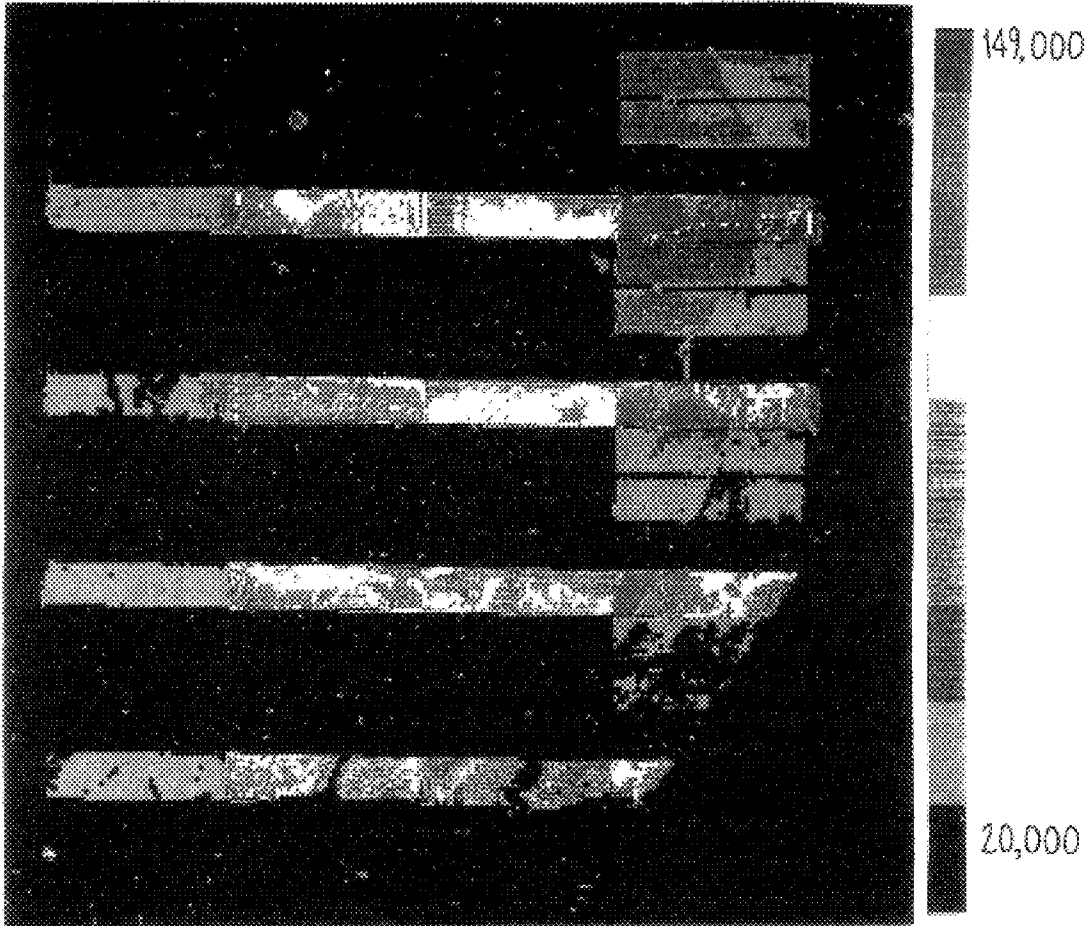


FIG. 23

U.S. Patent

Jun. 11, 2002

Sheet 27 of 27

US 6,403,320 B1

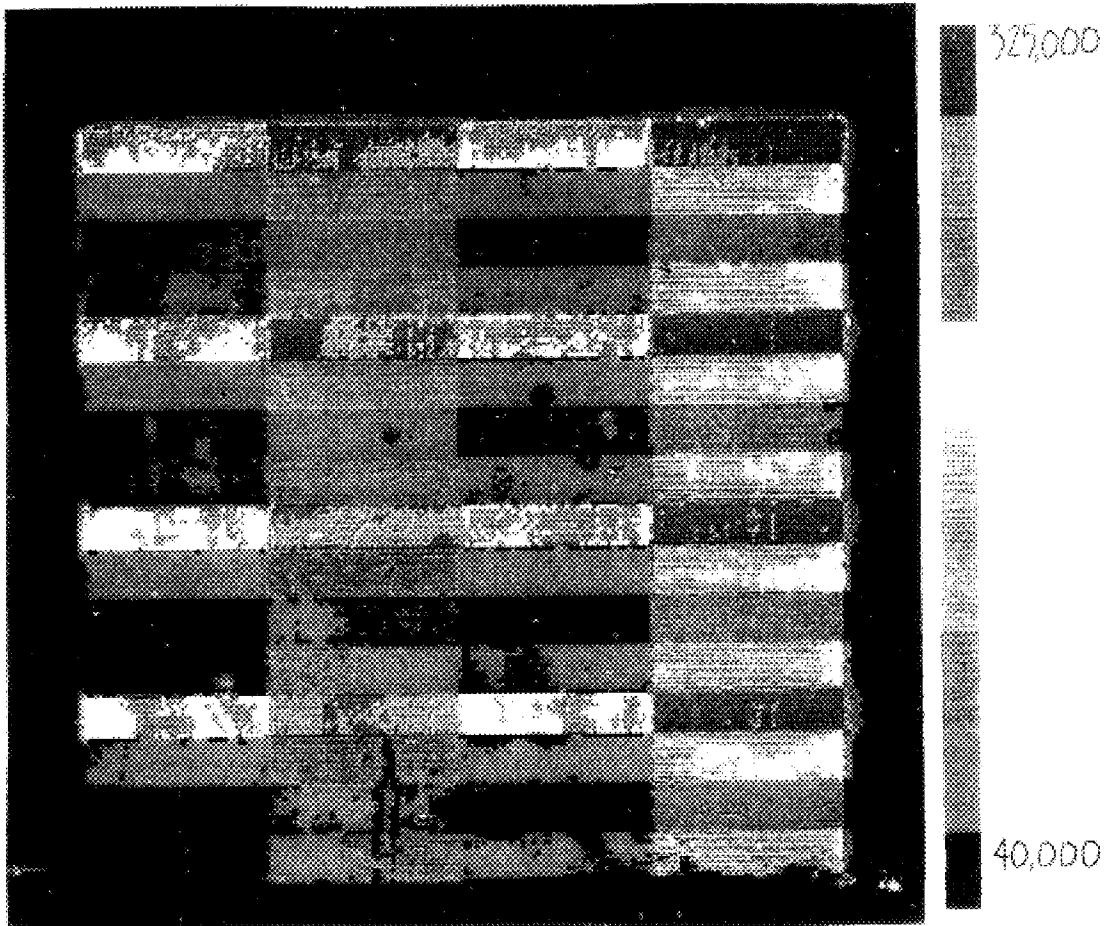


FIG. 24

US 6,403,320 B1

1

**SUPPORT BOUND PROBES AND METHODS
OF ANALYSIS USING THE SAME****CROSS-REFERENCE TO RELATED
APPLICATION**

The present application is a continuation of U.S. Ser. No. 09/557,875 filed Apr. 24, 2000, which is a continuation of U.S. Ser. No. 09/056,927 filed Apr. 8, 1998 now U.S. Pat. No. 6,197,506 which is a continuation of U.S. Ser. No. 08/670,118 filed Jun. 25, 1996, (now U.S. Pat. No. 5,800,992), which is a divisional of U.S. Ser. No. 08/168,904 filed Dec. 15, 1993 now abandoned, which is a continuation of U.S. Ser. No. 07/624,114, filed Dec. 6, 1990 (all incorporated by reference) now abandoned, which is a continuation-in-part of commonly assigned patent applications Pirrung et al., U.S. Ser. No. 07/362,901 (VLSIPS parent) filed on Jun. 7, 1989 now abandoned; and Pirrung et al., U.S. Ser. No. 07/492,462 (VLSIPS CIP), filed on Mar. 7, 1990 (now U.S. Pat. No. 5,143,854), which are hereby incorporated herein by reference. The present application is also a continuation-in-part of U.S. Ser. No. 08/348,471 filed Nov. 30, 1994, which is a continuation of U.S. Ser. No. 07/805,727 filed Dec. 6, 1991 (now U.S. Pat. No. 5,424,186), which is a continuation-in-part of U.S. Ser. No. 07/624,120, filed Dec. 6, 1990, which is a continuation-in-part of U.S. Ser. No. 07/492,462, filed Mar. 7, 1990 (now U.S. Pat. No. 5,143,854), which is a continuation-in-part of U.S. Ser. No. 07/362,901, filed Jun. 7, 1989 now abandoned. Additional commonly assigned applications Barrett et al., U.S. Ser. No. 07/435,316 (caged biotin parent) filed Nov. 13, 1989; and Barrett et al., U.S. Ser. No. 07/612,671 (caged biotin CIP), filed Nov. 13, 1990 are also incorporated herein by reference. Additional applications Pirrung et al., U.S. Ser. No. 07/624,120 (now abandoned) a divisional of which has issued as U.S. Pat. No. 5,744,101 and Dower et al., U.S. Ser. No. 07/626,730 (now U.S. Pat. No. 5,547,839), which are also commonly assigned and filed on the same day as this application, are also hereby incorporated herein by reference.

BACKGROUND OF THE INVENTION

The present invention relates to the sequencing, fingerprinting, and mapping of polymers, particularly biological polymers. The inventions may be applied, for example, in the sequencing, fingerprinting, or mapping of nucleic acids, polypeptides, oligosaccharides, and synthetic polymers.

The relationship between structure and function of macromolecules is of fundamental importance in the understanding of biological systems. These relationships are important to understanding, for example, the functions of enzymes, structural proteins, and signalling proteins, ways in which cells communicate with each other, as well as mechanisms of cellular control and metabolic feedback.

Genetic information is critical in continuation of life processes. Life is substantially informationally based and its genetic content controls the growth and reproduction of the organism and its complements. Polypeptides, which are critical features of all living systems, are encoded by the genetic material of the cell. In particular, the properties of enzymes, functional proteins, and structural proteins are determined by the sequence of amino acids which make them up. As structure and function are integrally related, many biological functions may be explained by elucidating the underlying the structural features which provide those functions. For this reason, it has become very important to

2

determine the genetic sequences of nucleotides which encode the enzymes, structural proteins, and other effectors of biological functions. In addition to segments of nucleotides which encode polypeptides, there are many nucleotide sequences which are involved in control and regulation of gene expression.

The human genome project is directed toward determining the complete sequence the genome of the human organism. Although such a sequence would not correspond to the sequence of any specific individual, it would provide significant information as to the general organization and specific sequences contained within segments from particular individuals. It would also provide mapping information which is very useful for further detailed studies. However, the need for highly rapid, accurate, and inexpensive sequencing technology is nowhere more apparent than in a demanding sequencing project such as this. To complete the sequencing of a human genome would require the determination of approximately 3×10^9 , or 3 billion base pairs.

The procedures typically used today for sequencing include the Sanger dideoxy method, see, e.g., Sanger et al. (1977) *Proc. Natl. Acad. Sci. USA*, 74:5463-5467, or the Maxam and Gilbert method, see, e.g., Maxam et al., (1980) *Methods in Enzymology*, 65:499-559. The Sanger method utilizes enzymatic elongation procedures with chain terminating nucleotides. The Maxam and Gilbert method uses chemical reactions exhibiting specificity of reaction to generate nucleotide specific cleavages. Both methods require a practitioner to perform a large number of complex manual manipulations. These manipulations usually require isolating homogeneous DNA fragments, elaborate and tedious preparing of samples, preparing a separating gel, applying samples to the gel, electrophoresing the samples into this gel, working up the finished gel, and analyzing the results of the procedure.

Thus, a less expensive, highly reliable, and labor efficient means for sequencing biological macromolecules is needed. A substantial reduction in cost and increase in speed of nucleotide sequencing would be very much welcomed. In particular, an automated system would improve the reproducibility and accuracy of procedures. The present invention satisfies these and other needs.

SUMMARY OF THE INVENTION

The present invention provides improved methods useful for de novo sequencing of an unknown polymer sequence, for verification of known sequences, for fingerprinting polymers, and for mapping homologous segments within a sequence. By reducing the number of manual manipulations required and automating most of the steps, the speed, accuracy, and reliability of these procedures are greatly enhanced.

The production of a substrate having a matrix of positionally defined regions with attached reagents exhibiting known recognition specificity can be used for the sequence analysis of a polymer. Although most directly applicable to sequencing, the present invention is also applicable to fingerprinting, mapping, and general screening of specific interactions. The VLSIPS substrates will be applied to evaluating other polymers, e.g., carbohydrates, polypeptides, hydrocarbon synthetic polymers, and the like. For these non-polynucleotides, the sequence specific reagents will usually be antibodies specific for a particular subunit sequence.

The present invention also provides a means to automate sequencing manipulations. The automation of the substrate

US 6,403,320 B1

3

production method and of the scan and analysis steps minimizes the need for human intervention. This simplifies the tasks and promotes reproducibility.

The present invention provides a composition comprising a plurality of positionally distinguishable sequence specific reagents attached to a solid substrate, which reagents are capable of specifically binding to a predetermined subunit sequence of a preselected multi-subunit length having at least three subunits, said reagents representing substantially all possible sequences of said preselected length. In some embodiments, the subunit sequence is a polynucleotide or a polypeptide, in others the preselected multi-subunit length is five subunits and the subunit sequence is a polynucleotide sequence. In other embodiments, the specific reagent is an oligonucleotide of at least about five nucleotides. Alternatively, the specific reagent is a monoclonal antibody. Usually the specific reagents are all attached to a single solid substrate, and the reagents comprise about 3000 different sequences. In other embodiments, the reagents represents at least about 25% of the possible subsequences of said preselected length. Usually, the reagents are localized in regions of the substrate having a density of at least 25 regions per square centimeter, and often the substrate has a surface area of less than about 4 square centimeters.

The present invention also provides methods for analyzing a sequence of a polynucleotide or a polypeptide, said method comprising the step of:

- a) exposing said polynucleotide or polypeptide to a composition as described.

It also provides useful methods for identifying or comparing a target sequence with a reference, said method comprising the step of:

- a) exposing said target sequence to a composition as described;
- b) determining the pattern of positions of the reagents which specifically interact with the target sequence; and
- c) comparing the pattern with the pattern exhibited by the reference when exposed to the composition.

The present invention also provides methods for sequencing a segment of a polynucleotide comprising the steps of:

- a) combining:
 - i) a substrate comprising a plurality of chemically synthesized and positionally distinguishable oligonucleotides capable of recognizing defined oligonucleotide sequences; and
 - ii) a target polynucleotide; thereby forming high fidelity matched duplex structures of complementary subsequences of known sequence; and
- b) determining which of said reagents have specifically interacted with subsequences in said target polynucleotide.

In one embodiment, the segment is substantially the entire length of said polynucleotide.

The invention also provides methods for sequencing a polymer, said method comprising the steps of:

- a) preparing a plurality of reagents which each specifically bind to a subsequence of preselected length;
- b) positionally attaching each of said reagents to one or more solid phase substrates, thereby producing substrates of positionally definable sequence specific probes;
- c) combining said substrates with a target polymer whose sequence is to be determined; and
- d) determining which of said reagents have specifically interacted with subsequences in said target polymer.

4

In one embodiment, the substrates are beads. Preferably, the plurality of reagents comprise substantially all possible subsequences of said preselected length found in said target. In another embodiment, the solid phase substrate is a single substrate having attached thereto reagents recognizing substantially all possible subsequences of preselected length found in said target.

In another embodiment, the method further comprises the step of analyzing a plurality of said recognized subsequences to assemble a sequence of said target polymer. In a bead embodiment, at least some of the plurality of substrates have one subsequence specific reagent attached thereto, and the substrates are coded to indicate the sequence specificity of said reagent.

The present invention also embraces a method of using a fluorescent nucleotide to detect interactions with oligonucleotide probes of known sequence, said method comprising:

- a) attaching said nucleotide to a target unknown polynucleotide sequence, and
- b) exposing said target polynucleotide sequence to a collection of positionally defined oligonucleotide probes of known sequences to determine the sequences of said probes which interact with said target.

In a further refinement, an additional step is included of:

- a) collating said known sequences to determine the overlaps of said known sequences to determine the sequence of said target sequence.

A method of mapping a plurality of sequences relative to one another is also provided, the method comprising:

- a) preparing a substrate having a plurality of positionally attached sequence specific probes are attached;
- b) exposing each of said sequences to said substrate, thereby determining the patterns of interaction between said sequence specific probes and said sequences; and
- c) determining the relative locations of said sequence specific probe interactions on said sequences to determine the overlaps and order of said sequences.

In one refinement, the sequence specific probes are oligonucleotides, applicable to where the target sequences are nucleic acid sequences.

In the nucleic acid sequencing application, the steps of the sequencing process comprise:

- a) producing a matrix substrate having known positionally defined regions of known sequence specific oligonucleotide probes;
- b) hybridizing a target polynucleotide to the positions on the matrix so that each of the positions which contain oligonucleotide probes complementary to a sequence on the target hybridize to the target molecule;
- c) detecting which positions have bound the target, thereby determining sequences which are found on the target; and
- d) analyzing the known sequences contained in the target to determine sequence overlaps and assembling the sequence of the target therefrom.

The enablement of the sequencing process by hybridization is based in large part upon the ability to synthesize a large number (e.g., to virtually saturate) of the possible overlapping sequence segments and distinguishing those probes which hybridize with fidelity from those which have mismatched bases, and to analyze a highly complex pattern of hybridization results to determine the overlap regions.

The detecting of the positions which bind the target sequence would typically be through a fluorescent label on the target. Although a fluorescent label is probably most

US 6,403,320 B1

5

convenient, other sorts of labels, e.g., radioactive, enzyme linked, optically detectable, or spectroscopic labels may be used. Because the oligonucleotide probes are positionally defined, the location of the hybridized duplex will directly translate to the sequences which hybridize. Thus, upon analysis of the positions provides a collection of subsequences found within the target sequence. These subsequences are matched with respect to their overlaps so as to assemble an intact target sequence.

In one preferred embodiment, linker molecules are provided on a substrate. A terminal end of the linker molecules is provided with a reactive functional group protected with a photoremovable protective group. Using lithographic methods, the photoremovable protective group is exposed to light and removed from the linker molecules in first selected regions. The substrate is then washed or otherwise contacted with a first monomer that reacts with exposed functional groups on the linker molecules. In a preferred embodiment, the monomer is an amino acid containing a photoremovable protective group at its amino or carboxy terminus and the linker molecule terminates in an amino or carboxy acid group bearing a photoremovable protective group.

A second set of selected regions is, thereafter, exposed to light and the photoremovable protective group on the linker molecule/protected amino acid is removed at the second set of regions. The substrate is then contacted with a second monomer containing a photoremovable protective group for reaction with exposed functional groups. This process is repeated to selectively apply monomers until polymers of a desired length and desired chemical sequence are obtained. Photolabile groups are then optionally removed and the sequence is, thereafter, optionally capped. Side chain protective groups, if present, are also removed.

An improved method and apparatus for the preparation of polymers is disclosed. The method and apparatus may be applied to synthesize a variety of polymers at known locations on a substrate. The method could be used to synthesize up to about 10^6 or more different sequences per cm^2 at known locations in some embodiments.

The method enables greater ease in peptide synthesis because the physical separation of reagents is not required when growing polymer chains. The chains themselves are separated by different physical locations on the substrate, but the entire substrate is exposed to the various reagents as the synthesis is conducted. Differential reaction is achieved by selectively exposing reactive functional groups to, e.g., light, electric currents, or another spatially localized activator. Remaining areas on the substrate remain unreacted.

By using the lithographic techniques disclosed herein, it is possible to direct light to relatively small and precisely known locations on the substrate. It is, therefore, possible to synthesize polymers of a known chemical sequence at known locations on the substrate.

The resulting substrate will have a variety of uses including, for example, screening large numbers of polymers for biological activity. To screen for biological activity, the substrate is exposed to one or more receptors such as antibody whole cells, receptors on vesicles, lipids, or any one of a variety of other receptors. The receptors are preferably labeled with, for example, a fluorescent marker, radioactive marker, or a labeled antibody reactive with the receptor. The location of the marker on the substrate is detected with, for example, photon detection or autoradiographic techniques. Through knowledge of the sequence of the material at the location where binding is detected, it is possible to quickly determine which sequence binds with the receptor and, therefore, the technique can be used to screen

6

large numbers of peptides. Other possible applications of the inventions herein include diagnostics in which various antibodies for particular receptors would be placed on a substrate and, for example, blood sera would be screened for immune deficiencies. Still further applications include, for example, selective "doping" of organic materials in semiconductor devices, and the like.

In connection with one aspect of the invention an improved reactor system for synthesizing polymers is also disclosed. The reactor system includes a substrate mount which engages a substrate around a periphery thereof. The substrate mount provides for a reactor space between the substrate and the mount through or into which reaction fluids are pumped or flowed. A mask is placed on or focused on the substrate and illuminated so as to deprotect selected regions of the substrate in the reactor space. A monomer is pumped through the reactor space or otherwise contacted with the substrate and reacts with the deprotected regions. By selectively deprotecting regions on the substrate and flowing predetermined monomers through the reactor space, desired polymers at known locations may be synthesized.

Improved detection apparatus and methods are also disclosed. The detection method and apparatus utilize a substrate having a large variety of polymer sequences at known locations on a surface thereof. The substrate is exposed to a fluorescently labeled receptor which binds to one or more of the polymer sequences. The substrate is placed in a microscope detection apparatus for identification of locations where binding takes place. The microscope detection apparatus includes a monochromatic or polychromatic light source for directing light at the substrate, means for detecting fluoresced light from the substrate, and means for determining a location of the fluoresced light. The means for detecting light fluoresced on the substrate may in some embodiments include a photon counter. The means for determining a location of the fluoresced light may include an x/y translation table for the substrate. Translation of the slide and data collection are recorded and managed by an appropriately programmed digital computer.

A further understanding of the nature and advantages of the inventions herein may be realized by reference to the remaining portions of the specification and the attached drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 illustrates a flow chart for sequence, fingerprint, or mapping analysis.

FIG. 2 illustrates the proper function of a VLSIPS peptide synthesis.

FIG. 3 illustrates the proper function of a VLSIPS dipeptide synthesis.

FIG. 4 illustrates the process of a VLSIPS trinucleotide synthesis.

FIG. 5 illustrates masking and irradiation of a substrate at a first location. The substrate is shown in cross-section;

FIG. 6 illustrates the substrate after application of a monomer "A";

FIG. 7 illustrates irradiation of the substrate at a second location;

FIG. 8 illustrates the substrate after application of monomer "B";

FIG. 9 illustrates irradiation of the "A" monomer;

FIG. 10 illustrates the substrate after a second application of "B";

FIG. 11 illustrates a completed substrate;

US 6,403,320 B1

7

FIGS. 12A and 12B illustrate alternative embodiments of a reactor system for forming a plurality of polymers on a substrate;

FIG. 13 illustrates a detection apparatus for locating fluorescent markers on the substrate;

FIGS. 14A–14M illustrate the method as it is applied to the production of the trimers of monomers “A” and “B”;

FIGS. 15A and 15B are fluorescence traces for standard fluorescent beads;

FIGS. 16A and 16B are fluorescence curves for NVOCl slides not exposed and exposed to light respectively;

FIGS. 17A to 17D are fluorescence plots of slides exposed through 100 μm , 50 μm , 20 μm , and 10 μm masks;

FIG. 18 illustrates fluorescence of a slide with the peptide YGGFL on selected regions of its surface which has been exposed to labeled Herz antibody specific for this sequence;

FIGS. 19A to 19D illustrate formation of and a fluorescence plot of a slide with a checkerboard pattern of YGGFL and GGFL exposed to labeled Herz antibody. FIG. 19C illustrates a 500 \times 500 μm mask which has been focused on the substrate according to FIG. 12A while FIG. 19D illustrates a 50 \times 50 μm mask placed in direct contact with the substrate in accord with FIG. 12B;

FIG. 20 is a fluorescence plot of YGGFL and PGGFL synthesized in a 50 μm checkerboard pattern;

FIG. 21 is a fluorescence plot of YPGGFL and YGGFL synthesized in a 50 μm checkerboard pattern;

FIGS. 22A and 22B illustrate the mapping of sixteen sequences synthesized on two different glass slides;

FIG. 23 is a fluorescence plot of the slide illustrated in FIG. 22A; and

FIG. 24 is a fluorescence plot of the slide illustrated in FIG. 14B.

GLOSSARY

The following terms are intended to have the following general meanings as they are used herein:

1. Complementary

Refers to the topological compatibility or matching together of interacting surfaces of a ligand molecule and its receptor. Thus, the receptor and its ligand can be described as complementary, and furthermore, the contact surface characteristics are complementary to each other.

2. Epitope

The portion of an antigen molecule which is delineated by the area of interaction with the subclass of receptors known as antibodies.

3. Ligand

A ligand is a molecule that is recognized by a particular receptor. Examples of ligands that can be investigated by this invention include, but are not restricted to, agonists and antagonists for cell membrane receptors, toxins and venoms, viral epitopes, hormones (e.g., opiates, steroids, etc.), hormone receptors, peptides, enzymes, enzyme substrates, cofactors, drugs, lectins, sugars, oligonucleotides, nucleic acids, oligosaccharides, proteins, and monoclonal antibodies.

4. Monomer

A member of the set of small molecules which can be joined together to form a polymer. The set of monomers includes but is not restricted to, for example, the set of common L-amino acids, the set of D-amino acids, the set of synthetic amino acids, the set of nucleotides and the set of pentoses and hexoses. As used herein, monomers refers to any member of a basis set for synthesis of a polymer. For

8

example, dimers of L-amino acids form a basis set of 400 monomers for synthesis of polypeptides. Different basis sets of monomers may be used at successive steps in the synthesis of a polymer.

5. Peptide

A polymer in which the monomers are alpha amino acids and which are joined together through amide bonds and alternatively referred to as a polypeptide. In the context of this specification it should be appreciated that the amino acids may be the L-optical isomer or the D-optical isomer. Peptides are more than two amino acid monomers long, and often more than 20 amino acid monomers long. Standard abbreviations for amino acids are used (e.g., P for proline). These abbreviations are included in Stryer, *Biochemistry*, Third Ed., 1988, which is incorporated herein by reference for all purposes.

6. Radiation

Energy which may be selectively applied including energy having a wavelength of between 10^{-14} and 10^4 meters including, for example, electron beam radiation, gamma radiation, x-ray radiation, ultra-violet radiation, visible light, infrared radiation, microwave radiation, and radio waves. “Irradiation” refers to the application of radiation to a surface.

7. Receptor

A molecule that has an affinity for a given ligand. Receptors may be naturally-occurring or manmade molecules. Also, they can be employed in their unaltered state or as aggregates with other species. Receptors may be attached, covalently or noncovalently, to a binding member, either directly or via a specific binding substance. Examples of receptors which can be employed by this invention include, but are not restricted to, antibodies, cell membrane receptors, monoclonal antibodies and antisera reactive with specific antigenic determinants (such as on viruses, cells or other materials), drugs, polynucleotides, nucleic acids, peptides, cofactors, lectins, sugars, polysaccharides, cells, cellular membranes, and organelles. Receptors are sometimes referred to in the art as anti-ligands. As the term receptors is used herein, no difference in meaning is intended. A “Ligand Receptor Pair” is formed when two macromolecules have combined through molecular recognition to form a complex.

Other examples of receptors which can be investigated by this invention include but are not restricted to:

a) Microorganism Receptors

Determination of ligands which bind to receptors, such as specific transport proteins or enzymes essential to survival of microorganisms, is useful in a new class of antibiotics. Of particular value would be antibiotics against opportunistic fungi, protozoa, and those bacteria resistant to the antibiotics in current use.

b) Enzymes

For instance, the binding site of enzymes such as the enzymes responsible for cleaving neurotransmitters; determination of ligands which bind to certain receptors to modulate the action of the enzymes which cleave the different neurotransmitters is useful in the development of drugs which can be used in the treatment of disorders of neurotransmission.

c) Antibodies

For instance, the invention may be useful in investigating the ligand-binding site on the antibody molecule which combines with the epitope of an antigen of interest; determining a sequence that mimics an antigenic epitope may lead to the development of vaccines of which the immunogen is based on one or more of such sequences or lead to the

US 6,403,320 B1

9

development of related diagnostic agents or compounds useful in therapeutic treatments such as for autoimmune diseases (e.g., by blocking the binding of the "self" antibodies).

d) Nucleic Acids

Sequences of nucleic acids may be synthesized to establish DNA or RNA binding sequences.

e) Catalytic Polypeptides

Polymers, preferably polypeptides, which are capable of promoting a chemical reaction involving the conversion of one or more reactants to one or more products. Such polypeptides generally include a binding site specific for at least one reactant or reaction intermediate and an active functionality proximate to the binding site, which functionality is capable of chemically modifying the bound reactant. Catalytic polypeptides are described in, for example, U.S. application Ser. No. 404,920, which is incorporated herein by reference for all purposes.

f) Hormone Receptors

For instance, the receptors for insulin and growth hormone. Determination of the ligands which bind with high affinity to a receptor is useful in the development of, for example, an oral replacement of the daily injections which diabetics must take to relieve the symptoms of diabetes, and in the other case, a replacement for the scarce human growth hormone which can only be obtained from cadavers or by recombinant DNA technology. Other examples are the vasoconstrictive hormone receptors; determination of those ligands which bind to a receptor may lead to the development of drugs to control blood pressure.

g) Opiate Receptors

Determination of ligands which bind to the opiate receptors in the brain is useful in the development of less-addictive replacements for morphine and related drugs.

8. Substrate

A material having a rigid or semi-rigid surface. In many embodiments, at least one surface of the substrate will be substantially flat, although in some embodiments it may be desirable to physically separate synthesis regions for different polymers with, for example, wells, raised regions, etched trenches, or the like. According to other embodiments, small beads may be provided on the surface which may be released upon completion of the synthesis.

9. Protective Group

A material which is bound to a monomer unit and which may be spatially removed upon selective exposure to an activator such as electromagnetic radiation. Examples of protective groups with utility herein include Nitroveratryloxy carbonyl, Nitrobenzyloxy carbonyl, Dimethyl dimethoxybenzyloxy carbonyl, 5-Bromo-7-nitroindoliny, o-Hydroxy- α -methyl cinnamoyl, and 2-Oxymethylene anthraquinone. Other examples of activators include ion beams, electric fields, magnetic fields, electron beams, x-ray, and the like.

10. Predefined Region

A predefined region is a localized area on a surface which is, was, or is intended to be activated for formation of a polymer. The predefined region may have any convenient shape, e.g., circular, rectangular, elliptical, wedge-shaped, etc. For the sake of brevity herein, "predefined regions" are sometimes referred to simply as "regions."

11. Substantially Pure

A polymer is considered to be "substantially pure" within a predefined region of a substrate when it exhibits characteristics that distinguish it from other predefined regions. Typically, purity will be measured in terms of biological activity or function as a result of uniform sequence. Such

10

characteristics will typically be measured by way of binding with a selected ligand or receptor.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

I. Overall Description

A. general

B. VLSIPS substrates

C. binary masking

D. applications

E. detection methods and apparatus

F. data analysis

II. Theoretical Analysis

A. simple n-mer structure; theory

15 B. complications

C. non-polynucleotide embodiments

III. Polynucleotide Sequencing

A. preparation of substrate matrix

B. labeling target polynucleotide

20 C. hybridization conditions

D. detection; VLSIPS scanning

E. analysis

F. substrate reuse

G. non-polynucleotide aspects

25 IV. Fingerprinting

A. general

B. preparation of substrate matrix

C. labeling target nucleotides

D. hybridization conditions

30 E. detection; VLSIPS scanning

F. analysis

G. substrate reuse

H. non-polynucleotide aspects

V. Mapping

35 A. general

B. preparation of substrate matrix

C. labeling

D. hybridization/specific interaction

40 E. detection

F. analysis

G. substrate reuse

H. non-polynucleotide aspects

VI. Additional Screening

A. specific interactions

45 B. sequence comparisons

C. categorizations

D. statistical correlations

VII. Formation of Substrate

A. instrumentation

B. binary masking

C. synthetic methods

D. surface immobilization

VIII. Hybridization/Specific Interaction

55 A. general

B. important parameters

IX. Detection Methods

A. labeling techniques

B. scanning system

60 X. Data Analysis

A. general

B. hardware

C. software

XI. Substrate Reuse

65 A. removal of label

B. storage and preservation

C. processes to avoid degradation of oligomers

US 6,403,320 B1

11

XII. Integrated Sequencing Strategy

- A. initial mapping strategy
- B. selection of smaller clones
- C. actual sequencing procedures

XIII. Commercial Applications

- A. sequencing
- B. fingerprinting
- C. mapping

I. Overall Description

A. General

The present invention relies in part on the ability to synthesize or attach specific recognition reagents at known locations on a substrate, typically a single substrate. In particular, the present invention provides the ability to prepare a substrate having a very high density matrix pattern of positionally defined specific recognition reagents. The reagents are capable of interacting with their specific targets while attached to the substrate, e.g., solid phase interactions, and by appropriate labeling of these targets, the sites of the interactions between the target and the specific reagents may be derived. Because the reagents are positionally defined, the sites of the interactions will define the specificity of each interaction. As a result, a map of the patterns of interactions with specific reagents on the substrate is convertible into information on the specific interactions taking place, e.g., the recognized features. Where the specific reagents recognize a large number of possible features, this system allows the determination of the combination of specific interactions which exist on the target molecule. Where the number of features is sufficiently large, the identical same combination, or pattern, of features is sufficiently unlikely that a particular target molecule may often be uniquely defined by its features. In the extreme, the features may actually be the subunit sequence of the target molecule, and a given target sequence may be uniquely defined by its combination of features.

In particular, the methodology is applicable to sequencing polynucleotides. The specific sequence recognition reagents will typically be oligonucleotide probes which hybridize with specificity to subsequences found on the target sequence. A sufficiently large number of those probes allows the fingerprinting of a target polynucleotide or the relative mapping of a collection of target polynucleotides, as described in greater detail below.

In the high resolution fingerprinting provided by a saturating collection of probes which include all possible subsequences of a given size, e.g., 10-mers, collating of all the subsequences and determination of specific overlaps will be derived and the entire sequence can usually be reconstructed.

Although a polynucleotide sequence analysis is a preferred embodiment, for which the specific reagents are most easily accessible, the invention is also applicable to analysis of other polymers, including polypeptides, carbohydrates, and synthetic polymers, including α -, β -, and ω -amino acids, polyurethanes, polyesters, polycarbonates, polyureas, polyamides, polyethyleneimines, polyarylene sulfides, polysiloxanes, polyimides, polyacetates, and mixed polymers. Various optical isomers, e.g., various D- and L-forms of the monomers, may be used.

Sequence analysis will take the form of complete sequence determination, to the level of the sequence of individual subunits along the entire length of the target sequence. Sequence analysis also takes the form of sequence homology, e.g., less than absolute subunit resolution, where "similarity" in the sequence will be detectable, or the form of selective sequences of homology interspersed at specific or irregular locations.

12

In either case, the sequence is determinable at selective resolution or at particular locations. Thus, the hybridization method will be useful as a means for identification, e.g., a "fingerprint", much like a Southern hybridization method is used. It is also useful to map particular target sequences.

B. VLSIPS Substrates

The invention is enabled by the development of technology to prepare substrates on which specific reagents may be either positionally attached or synthesized. In particular, the very large scale immobilized polymer synthesis (VLSIPS) technology allows for the very high density production of an enormous diversity of reagents mapped out in a known matrix pattern on a substrate. These reagents specifically recognize subsequences in a target polymer and bind thereto, producing a map of positionally defined regions of interaction. These map positions are convertible into actual features recognized, and thus would be present in the target molecule of interest.

As indicated, the sequence specific recognition reagents will often be oligonucleotides which hybridize with fidelity and discrimination to the target sequence. For use with other polymers, monoclonal or polyclonal antibodies having high sequence specificity will often be used.

In the generic sense, the VLSIPS technology allows the production of a substrate with a high density matrix of positionally mapped regions with specific recognition reagents attached at each distinct region. By use of protective groups which can be positionally removed, or added, the regions can be activated or deactivated for addition of particular reagents or compounds. Details of the protection are described below and in related application U.S. Ser. No. 07/492,462 (VLSIPS CIP). In a preferred embodiment, photosensitive protecting agents will be used and the regions of activation or deactivation may be controlled by electro-optical and optical methods, similar to many of the processes used in semiconductor wafer and chip fabrication.

In the nucleic acid nucleotide sequencing application, a VLSIPS substrate is synthesized having positionally defined oligonucleotide probes. See U.S. Ser. No. 07/492,462 (VLSIPS CIP); and U.S. Ser. No. 07/624,120, (automated VLSIPS). By use of masking technology and photosensitive synthetic subunits, the VLSIPS apparatus allows for the stepwise synthesis of polymers according to a positionally defined matrix pattern. Each oligonucleotide probe will be synthesized at known and defined positional locations on the substrate. This forms a matrix pattern of known relationship between position and specificity of interaction. The VLSIPS technology allows the production of a very large number of different oligonucleotide probes to be simultaneously and automatically synthesized including numbers in excess of about 10^2 , 10^3 , 10^4 , 10^5 , 10^6 , or even more, and at densities of at least about 10^2 , $10^3/\text{cm}^2$, $10^4/\text{cm}^2$, $10^5/\text{cm}^2$ and up to $10^6/\text{cm}^2$ or more. This application discloses methods for synthesizing polymers on a silicon or other suitably derivatized substrate, methods and chemistry for synthesizing specific types of biological polymers on those substrates, apparatus for scanning and detecting whether interaction has occurred at specific locations on the substrate, and various other technologies related to the use of a high density very large scale immobilized polymer substrate. In particular, sequencing, fingerprinting, and mapping applications are discussed herein in detail, though related technologies are described in simultaneously filed applications U.S. Ser. No. 07/624,1200 (automated VLSIPS) and U.S. Ser. No. 07/626,730, (sequencing by synthesis), each of which is hereby incorporated herein by reference.

In other embodiments, antibody probes will be generated which specifically recognize particular subsequences found

US 6,403,320 B1

13

on a polymer. Antibodies would be generated which are specific for recognizing a three contiguous amino acid sequence, and monoclonal antibodies may be preferred. Optimally, these antibodies would not recognize any sequences other than the specific three amino acid stretch desired and the binding affinity should be insensitive to flanking or remote sequences found on a target molecule. Likewise, antibodies specific for particular carbohydrate linkages or sequences will be generated. A similar approach could be used for preparing specific reagents which recognize other polymer subunit sequences. These reagents would typically be site specifically localized to a substrate matrix pattern where the regions are closely packed.

These reagents could be individually attached at specific sites on the substrate in a matrix by an automated procedure where the regions are positionally targeted by some other specific mechanism, e.g., one which would allow the entire collection of reagents to be attached to the substrate in a single reaction. Each reagent could be separately attached to a specific oligonucleotide sequence by an automated procedure. This would produce a collection of reagents where, e.g., each monoclonal antibody would have a unique oligonucleotide sequence attached to it. By virtue of a VLSIPS substrate which has different complementary oligonucleotides synthesized on it, each monoclonal antibody would specifically be bound only at that site on the substrate where the complementary oligonucleotide has been synthesized. A crosslinking step would fix the reagent to the substrate. See, e.g., Dattagupta et al. (1985) U.S. Pat. No. 4,542,102 and (1987) U.S. Pat. No. 4,713,326; and Chatterjee, M. et al. (1990) *J. Am. Chem. Soc.* 112:6397-6399, which are hereby incorporated herein by reference. This allows a high density positionally specific collection of specific recognition reagents, e.g., monoclonal antibodies, to be immobilized to a solid substrate using an automated system.

The regions which define particular reagents will usually be generated by selective protecting groups which may be activated or deactivated. Typically the protecting group will be bound to a monomer subunit or spatial region, and can be spatially affected by an activator, such as electromagnetic radiation. Examples of protective groups with utility herein include nitroveratryl oxycarbonyl (NVOC), nitrobenzyl oxycarbonyl (NBOC), dimethyl dimethoxy benzyloxy carbonyl, 5-bromo-7-nitroindolyl, O-hydroxy- α -methyl cinnamoyl, and 2-oxymethylene anthraquinone. Examples of activators include ion beams, electric fields, magnetic fields, electron beams, x-ray, and other forms of electromagnetic radiation.

The present invention provides methods and apparatus for the preparation and use of a substrate having a plurality of polymer sequences in predefined regions. The invention is described herein primarily with regard to the preparation of molecules containing sequences of amino acids, but could readily be applied in the preparation of other polymers. Such polymers include, for example, both linear and cyclic polymers of nucleic acids, polysaccharides, phospholipids, and peptides having either α -, β -, or ω -amino acids, heteropolymers in which a known drug is covalently bound to any of the above, polyurethanes, polyesters, polycarbonates, polyureas, polyamides, polyethyleneimines, polyarylene sulfides, polysiloxanes, polyimides, polyacetates, or other polymers which will be apparent upon review of this disclosure. In a preferred embodiment, the invention herein is used in the synthesis of peptides.

The prepared substrate may, for example, be used in screening a variety of polymers as ligands for binding with a receptor, although it will be apparent that the invention

14

could be used for the synthesis of a receptor for binding with a ligand. The substrate disclosed herein will have a wide variety of other uses. Merely by way of example, the invention herein can be used in determining peptide and nucleic acid sequences which bind to proteins, finding sequence-specific binding drugs, identifying epitopes recognized by antibodies, and evaluation of a variety of drugs for clinical and diagnostic applications, as well as combinations of the above.

The invention preferably provides for the use of a substrate "S" with a surface. Linker molecules "L" are optionally provided on a surface of the substrate. The purpose of the linker molecules, in some embodiments, is to facilitate receptor recognition of the synthesized polymers.

Optionally, the linker molecules may be chemically protected for storage purposes. A chemical storage protective group such as t-BOC (t-butoxycarbonyl) may be used in some embodiments. Such chemical protective groups would be chemically removed upon exposure to, for example, acidic solution and would serve to protect the surface during storage and be removed prior to polymer preparation.

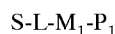
On the substrate or a distal end of the linker molecules, a functional group with a protective group P_0 is provided. The protective group P_0 may be removed upon exposure to radiation, electric fields, electric currents, or other activators to expose the functional group.

In a preferred embodiment, the radiation is ultraviolet (UV), infrared (IR), or visible light. As more fully described below, the protective group may alternatively be an electrochemically-sensitive group which may be removed in the presence of an electric field. In still further alternative embodiments, ion beams, electron beams, or the like may be used for deprotection.

In some embodiments, the exposed regions and, therefore, the area upon which each distinct polymer sequence is synthesized are smaller than about 1 cm^2 or less than 1 mm^2 . In preferred embodiments the exposed area is less than about $10,000\text{ }\mu\text{m}^2$ or, more preferably, less than $100\text{ }\mu\text{m}^2$ and may, in some embodiments, encompass the binding site for as few as a single molecule. Within these regions, each polymer is preferably synthesized in a substantially pure form.

Concurrently or after exposure of a known region of the substrate to light, the surface is contacted with a first monomer unit M_1 which reacts with the functional group which has been exposed by the deprotection step. The first monomer includes a protective group P_1 . P_1 may or may not be the same as P_0 .

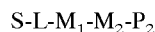
Accordingly, after a first cycle, known first regions of the surface may comprise the sequence:



while remaining regions of the surface comprise the sequence:



Thereafter, second regions of the surface (which may include the first region) are exposed to light and contacted with a second monomer M_2 (which may or may not be the same as M_1) having a protective group P_2 . P_2 may or may not be the same as P_0 and P_1 . After this second cycle, different regions of the substrate may comprise one or more of the following sequences:



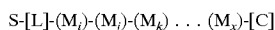
US 6,403,320 B1

15

S-L-M₁-P₁ and/or
S-L-P₀.

The above process is repeated until the substrate includes desired polymers of desired lengths. By controlling the locations of the substrate exposed to light and the reagents exposed to the substrate following exposure, the location of each sequence will be known.

Thereafter, the protective groups are removed from some or all of the substrate and the sequences are, optionally, capped with a capping unit C. The process results in a substrate having a surface with a plurality of polymers of the following general formula:



where square brackets indicate optional groups, and M_i . . . M_x indicates any sequence of monomers. The number of monomers could cover a wide variety of values, but in a preferred embodiment they will range from 2 to 100.

In some embodiments a plurality of locations on the substrate polymers are to contain a common monomer subsequence. For example, it may be desired to synthesize a sequence S-M₁-M₂-M₃ at first locations and a sequence S-M₄-M₂-M₃ at second locations. The process would commence with irradiation of the first locations followed by contacting with M₁-P, resulting in the sequence S-M₁-P at the first location. The second locations would then be irradiated and contacted with M₄-P, resulting in the sequence S-M₄-P at the second locations. Thereafter both the first and second locations would be irradiated and contacted with the dimer M₂-M₃, resulting in the sequence S-M₁-M₂-M₃ at the first locations and S-M₄-M₂-M₃ at the second locations. Of course, common subsequences of any length could be utilized including those in a range of 2 or more monomers, 2 to 100 monomers, 2 to 20 monomers, and a most preferred range of 2 to 3 monomers.

According to other embodiments, a set of masks is used for the first monomer layer and, thereafter, varied light wavelengths are used for selective deprotection. For example, in the process discussed above, first regions are first exposed through a mask and reacted with a first monomer having a first protective group P₁, which is removable upon exposure to a first wavelength of light (e.g., IR). Second regions are masked and reacted with a second monomer having a second protective group P₂, which is removable upon exposure to a second wavelength of light (e.g., UV). Thereafter, masks become unnecessary in the synthesis because the entire substrate may be exposed alternatively to the first and second wavelengths of light in the deprotection cycle.

The polymers prepared on a substrate according to the above methods will have a variety of uses including, for example, screening for biological activity. In such screening activities, the substrate containing the sequences is exposed to an unlabeled or labeled receptor such as an antibody, receptor on a cell, phospholipid vesicle, or any one of a variety of other receptors. In one preferred embodiment the polymers are exposed to a first, unlabeled receptor of interest and, thereafter, exposed to a labeled receptor-specific recognition element, which is, for example, an antibody. This process will provide signal amplification in the detection stage.

The receptor molecules may bind with one or more polymers on the substrate. The presence of the labeled receptor and, therefore, the presence of a sequence which binds with the receptor is detected in a preferred embodiment through the use of autoradiography, detection of fluo-

16

rescence with a charge-coupled device, fluorescence microscopy, or the like. The sequence of the polymer at the locations where the receptor binding is detected may be used to determine all or part of a sequence which is complementary to the receptor.

Use of the invention herein is illustrated primarily with reference to screening for biological activity. The invention will, however, find many other uses. For example, the invention may be used in information storage (e.g., on optical disks), production of molecular electronic devices, production of stationary phases in separation sciences, production of dyes and brightening agents, photography, and in immobilization of cells, proteins, lectins, nucleic acids, polysaccharides and the like in patterns on a surface via molecular recognition of specific polymer sequences. By synthesizing the same compound in adjacent, progressively differing concentrations, a gradient will be established to control chemotaxis or to develop diagnostic dipsticks which, for example, titrate an antibody against an increasing amount of antigen. By synthesizing several catalyst molecules in close proximity, more efficient multistep conversions may be achieved by "coordinate immobilization." Coordinate immobilization also may be used for electron transfer systems, as well as to provide both structural integrity and other desirable properties to materials such as lubrication, wetting, etc.

According to alternative embodiments, molecular biodistribution or pharmacokinetic properties may be examined. For example, to assess resistance to intestinal or serum proteases, polymers may be capped with a fluorescent tag and exposed to biological fluids of interest.

III. Polymer Synthesis

FIG. 1 illustrates one embodiment of the invention disclosed herein in which a substrate 2 is shown in cross-section. Essentially, any conceivable substrate may be employed in the invention. The substrate may be biological, nonbiological, organic, inorganic, or a combination of any of these, existing as particles, strands, precipitates, gels, sheets, tubing, spheres, containers, capillaries, pads, slices, films, plates, slides, etc. The substrate may have any convenient shape, such as a disc, square, sphere, circle, etc. The substrate is preferably flat but may take on a variety of alternative surface configurations. For example, the substrate may contain raised or depressed regions on which the synthesis takes place. The substrate and its surface preferably form a rigid support on which to carry out the reactions described herein. The substrate and its surface is also chosen to provide appropriate light-absorbing characteristics. For instance, the substrate may be a polymerized Langmuir Blodgett film, functionalized glass, Si, Ge, GaAs, GaP, SiO₂, SiN₄, modified silicon, or any one of a wide variety of gels or polymers such as (poly)tetrafluoro-ethylene, (poly)vinylidenedifluoride, polystyrene, polycarbonate, or combinations thereof. Other substrate materials will be readily apparent to those of skill in the art upon review of this disclosure. In a preferred embodiment the substrate is flat glass or single-crystal silicon with surface relief features of less than 10 Å.

According to some embodiments, the surface of the substrate is etched using well known techniques to provide for desired surface features. For example, by way of the formation of trenches, v-grooves, mesa structures, or the like, the synthesis regions may be more closely placed within the focus point of impinging light, be provided with reflective "mirror" structures for maximization of light collection from fluorescent sources, or the like.

Surfaces on the solid substrate will usually, though not always, be composed of the same material as the substrate.

US 6,403,320 B1

17

Thus, the surface may be composed of any of a wide variety of materials, for example, polymers, plastics, resins, polysaccharides, silica or silica-based materials, carbon, metals, inorganic glasses, membranes, or any of the above-listed substrate materials. In some embodiments the surface may provide for the use of caged binding members which are attached firmly to the surface of the substrate in accord with the teaching of copending application Ser. No. 404,920, previously incorporated herein by reference. Preferably, the surface will contain reactive groups, which could be carboxyl, amino, hydroxyl, or the like. Most preferably, the surface will be optically transparent and will have surface Si—OH functionalities, such as are found on silica surfaces.

The surface 4 of the substrate is preferably provided with a layer of linker molecules 6, although it will be understood that the linker molecules are not required elements of the invention. The linker molecules are preferably of sufficient length to permit polymers in a completed substrate to interact freely with molecules exposed to the substrate. The linker molecules should be 6–50 atoms long to provide sufficient exposure. The linker molecules may be, for example, aryl acetylene, ethylene glycol oligomers containing 2–10 monomer units, diamines, diacids, amino acids, or combinations thereof. Other linker molecules may be used in light of this disclosure.

According to alternative embodiments, the linker molecules are selected based upon their hydrophilic/hydrophobic properties to improve presentation of synthesized polymers to certain receptors. For example, in the case of a hydrophilic receptor, hydrophilic linker molecules will be preferred so as to permit the receptor to more closely approach the synthesized polymer.

According to another alternative embodiment, linker molecules are also provided with a photocleavable group at an intermediate position. The photocleavable group is preferably cleavable at a wavelength different from the protective group. This enables removal of the various polymers following completion of the synthesis by way of exposure to the different wavelengths of light.

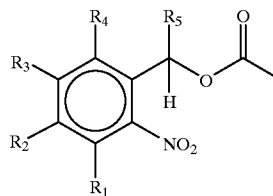
The linker molecules can be attached to the substrate via carbon-carbon bonds using, for example, (poly) trifluorochloroethylene surfaces, or preferably, by siloxane bonds (using, for example, glass or silicon oxide surfaces). Siloxane bonds with the surface of the substrate may be formed in one embodiment via reactions of linker molecules bearing trichlorosilyl groups. The linker molecules may optionally be attached in an ordered array, i.e., as parts of the head groups in a polymerized Langmuir Blodgett film. In alternative embodiments, the linker molecules are adsorbed to the surface of the substrate.

The linker molecules and monomers used herein are provided with a functional group to which is bound a protective group. Preferably, the protective group is on the distal or terminal end of the linker molecule opposite the substrate. The protective group may be either a negative protective group (i.e., the protective group renders the linker molecules less reactive with a monomer upon exposure) or a positive protective group (i.e., the protective group renders the linker molecules more reactive with a monomer upon exposure). In the case of negative protective groups an additional step of reactivation will be required. In some embodiments, this will be done by heating.

The protective group on the linker molecules may be selected from a wide variety of positive light-reactive groups preferably including nitro aromatic compounds such as o-nitrobenzyl derivatives or benzylsulfonfyl. In a preferred embodiment, 6-nitroveratryloxy-carbonyl (NVOC),

18

2-nitrobenzyloxycarbonyl (NBOC) or α,α -dimethyldimethoxybenzyloxycarbonyl (DDZ) is used. In one embodiment, a nitro aromatic compound containing a benzylic hydrogen ortho to the nitro group is used, i.e., a chemical of the form:



where R₁ is alkoxy, alkyl, halo, aryl, alkenyl, or hydrogen; R₂ is alkoxy, alkyl, halo, aryl, nitro, or hydrogen; R₃ is alkoxy, alkyl, halo, nitro, aryl, or hydrogen; R₄ is alkoxy, alkyl, hydrogen, aryl, halo, or nitro; and R₅ is alkyl, alkynyl, cyano, alkoxy, hydrogen, halo, aryl, or alkenyl. Other materials which may be used include o-hydroxy- α -methyl cinnamoyl derivatives. Photoremovable protective groups are described in, for example, Patchornik, *J. Am. Chem. Soc.* (1970) 92:6333 and Amit et al., *J. Org. Chem.* (1974) 39:192, both of which are incorporated herein by reference.

In an alternative embodiment the positive reactive group is activated for reaction with reagents in solution. For example, a 5-bromo-7-nitro indoline group, when bound to a carbonyl, undergoes reaction upon exposure to light at 420 nm.

In a second alternative embodiment, the reactive group on the linker molecule is selected from a wide variety of negative light-reactive groups including a cinnamate group.

Alternatively, the reactive group is activated or deactivated by electron beam lithography, x-ray lithography, or any other radiation. Suitable reactive groups for electron beam lithography include sulfonfyl. Other methods may be used including, for example, exposure to a current source. Other reactive groups and methods of activation may be used in light of this disclosure.

As shown in FIG. 5, the linking molecules are preferably exposed to, for example, light through a suitable mask 8 using photolithographic techniques of the type known in the semiconductor industry and described in, for example, Sze, *VLSI Technology*, McGraw-Hill (1983), and Mead et al., *Introduction to VLSI Systems*, Addison-Wesley (1980), which are incorporated herein by reference for all purposes. The light may be directed at either the surface containing the protective groups or at the back of the-substrate, so long as the substrate is transparent to the wavelength of light needed for removal of the protective groups. In the embodiment shown in FIG. 5, light is directed at the surface of the substrate containing the protective groups. FIG. 5 illustrates the use of such masking techniques as they are applied to a positive reactive group so as to activate linking molecules and expose functional groups in areas 10a and 10b.

The mask 8 is in one embodiment a transparent support material selectively coated with a layer of opaque material. Portions of the opaque material are removed, leaving opaque material in the precise pattern desired on the substrate surface. The mask is brought into close proximity with, imaged on, or brought directly into contact with the substrate surface as shown in FIG. 5. "Openings" in the mask correspond to locations on the substrate where it is desired to remove photoremovable protective groups from the sub-

US 6,403,320 B1

19

strate. Alignment may be performed using conventional alignment techniques in which alignment marks (not shown) are used to accurately overlay successive masks with previous patterning steps, or more sophisticated techniques may be used. For example, interferometric techniques such as the one described in Flanders et al., "A New Interferometric Alignment Technique," *App. Phys. Lett.* (1977) 31:426-428, which is incorporated herein by reference, may be used.

To enhance contrast of light applied to the substrate, it is desirable to provide contrast enhancement materials between the mask and the substrate according to some embodiments. This contrast enhancement layer may comprise a molecule which is decomposed by light such as guinone diazid or a material which is transiently bleached at the wavelength of interest. Transient bleaching of materials will allow greater penetration where light is applied, thereby enhancing contrast. Alternatively, contrast enhancement may be provided by way of a cladded fiber optic bundle.

The light may be from a conventional incandescent source, a laser, a laser diode, or the like. If non-collimated sources of light are used it may be desirable to provide a thick- or multi-layered mask to prevent spreading of the light onto the substrate. It may, further, be desirable in some embodiments to utilize groups which are sensitive to different wavelengths to control synthesis. For example, by using groups which are sensitive to different wavelengths, it is possible to select branch positions in the synthesis of a polymer or eliminate certain masking steps. Several reactive groups along with their corresponding wavelengths for deprotection are provided in Table 1.

TABLE 1

Group	Approximate Deprotection Wavelength
Nitroveratryloxy carbonyl (NVOC)	UV (300-400 nm)
Nitrobenzyloxy carbonyl (NBOC)	UV (300-350 nm)
Dimethyl dimethoxybenzyloxy carbonyl	UV (280-300 nm)
5-Bromo-7-nitroindoliny	UV (420 nm)
o-Hydroxy- α -methyl cinnamoyl	UV (300-350 nm)
2-Oxymethylene anthraquinone	UV (350 nm)

While the invention is illustrated primarily herein by way of the use of a mask to illuminate selected regions the substrate, other techniques may also be used. For example, the substrate may be translated under a modulated laser or diode light source. Such techniques are discussed in, for example, U.S. Pat. No. 4,719,615 (Feyrer et al.), which is incorporated herein by reference. In alternative embodiments a laser galvanometric scanner is utilized. In other embodiments, the synthesis may take place on or in contact with a conventional liquid crystal (referred to herein as a "light valve") or fiber optic light sources. By appropriately modulating liquid crystals, light may be selectively controlled so as to permit light to contact selected regions of the substrate. Alternatively, synthesis may take place on the end of a series of optical fibers to which light is selectively applied. Other means of controlling the location of light exposure will be apparent to those of skill in the art.

The substrate may be irradiated either in contact or not in contact with a solution (not shown) and is, preferably, irradiated in contact with a solution. The solution contains reagents to prevent the by-products formed by irradiation from interfering with synthesis of the polymer according to some embodiments. Such by-products might include, for example, carbon dioxide, nitrosocarbonyl compounds, styrene derivatives, indole derivatives, and products of their photochemical reactions. Alternatively, the solution may

20

contain reagents used to match the index of refraction of the substrate. Reagents added to the solution may further include, for example, acidic or basic buffers, thiols, substituted hydrazines and hydroxylamines, reducing agents (e.g., NADH) or reagents known to react with a given functional group (e.g., aryl nitroso+glyoxylic acid \rightarrow aryl formhydroxamate+CO₂).

Either concurrently with or after the irradiation step, the linker molecules are washed or otherwise contacted with a first monomer, illustrated by "A" in regions 12a and 12b in FIG. 6. The first monomer reacts with the activated functional groups of the linkage molecules which have been exposed to light. The first monomer, which is preferably an amino acid, is also provided with a photoprotective group. The photoprotective group on the monomer may be the same as or different than the protective group used in the linkage molecules, and may be selected from any of the above-described protective groups. In one embodiment, the protective groups for the A monomer is selected from the group NBOC and NVOC.

As shown in FIG. 7 the process of irradiating is thereafter repeated, with a mask repositioned so as to remove linkage protective groups and expose functional groups in regions 14a and 14b which are illustrated as being regions which were protected in the previous masking step. As an alternative to repositioning of the first mask, in many embodiments a second mask will be utilized. In other alternative embodiments, some steps may provide for illuminating a common region in successive steps. As shown in FIG. 7, it may be desirable to provide separation between irradiated regions. For example, separation of about 1-5 μ m may be appropriate to account for alignment tolerances.

As shown in FIG. 8, the substrate is then exposed to a second protected monomer "B," producing B regions 16a and 16b. Thereafter, the substrate is again masked so as to remove the protective groups and expose reactive groups on A region 12a and B region 16b. The substrate is again exposed to monomer B, resulting in the formation of the structure shown in FIG. 10. The dimers B-A and B-B have been produced on the substrate.

A subsequent series of masking and contacting steps similar to those described above with A (not shown) provides the structure shown in FIG. 11. The process provides all possible dimers of B and A, i.e., B-A, A-B, A-A, and B-B.

The substrate, the area of synthesis, and the area for synthesis of each individual polymer could be of any size or shape. For example, squares, ellipsoids, rectangles, triangles, circles, or portions thereof, along with irregular geometric shapes, may be utilized. Duplicate synthesis areas may also be applied to a single substrate for purposes of redundancy.

In one embodiment the regions 12 and 16 on the substrate will have a surface area of between about 1 cm² and 10⁻¹⁰ cm². In some embodiments the regions 12 and 16 have areas of less than about 10⁻¹⁰ cm², 10⁻² cm², 10⁻³ cm², 10⁻⁴ cm², 10⁻⁵ cm², 10⁻⁶ cm², 10⁻⁷ cm², 10⁻⁸ cm², or 10⁻¹⁰ cm². In a preferred embodiment, the regions 12 and 16 are between about 10 \times 10 μ m and 500 \times 500 μ m.

In some embodiments a single substrate supports more than about 10 different monomer sequences and preferably more than about 100 different monomer sequences, although in some embodiments more than about 10³, 10⁴, 10⁵, 10⁶, 10⁷, or 10⁸ different sequences are provided on a substrate. Of course, within a region of the substrate in which a monomer sequence is synthesized, it is preferred that the monomer sequence be substantially pure. In some embodiments, regions of the substrate contain polymer

US 6,403,320 B1

21

sequences which are at least about 1%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, or 99% pure.

According to some embodiments, several sequences are intentionally provided within a single region so as to provide an initial screening for biological activity, after which materials within regions exhibiting significant binding are further evaluated.

IV. Details of One Embodiment of a Reactor System

FIG. 12A schematically illustrates a preferred embodiment of a reactor system **100** for synthesizing polymers on the prepared substrate in accordance with one aspect of the invention. The reactor system includes a body **102** with a cavity **104** on a surface thereof. In preferred embodiments the cavity **104** is between about 50 and 1000 μm deep with a depth of about 500 μm preferred.

The bottom of the cavity is preferably provided with an array of ridges **106** which extend both into the plane of the Figure and parallel to the plane of the Figure. The ridges are preferably about 50 to 200 μm deep and spaced at about 2 to 3 mm. The purpose of the ridges is to generate turbulent flow for better mixing. The bottom surface of the cavity is preferably light absorbing so as to prevent reflection of impinging light.

A substrate **112** is mounted above the cavity **104**. The substrate is provided along its bottom surface **114** with a photoremovable protective group such as NVOC with or without an intervening linker molecule. The substrate is preferably transparent to a wide spectrum of light, but in some embodiments is transparent only at a wavelength at which the protective group may be removed (such as UV in the case of NVOC). The substrate in some embodiments is a conventional microscope glass slide or cover slip. The substrate is preferably as thin as possible, while still providing adequate physical support. Preferably, the substrate is less than about 1 mm thick, more preferably less than 0.5 mm thick, more preferably less than 0.1 mm thick, and most preferably less than 0.05 mm thick. In alternative preferred embodiments, the substrate is quartz or silicon.

The substrate and the body serve to seal the cavity except for an inlet port **108** and an outlet port **110**. The body and the substrate may be mated for sealing in some embodiments with one or more gaskets. According to a preferred embodiment, the body is provided with two concentric gaskets and the intervening space is held at vacuum to ensure mating of the substrate to the gaskets.

Fluid is pumped through the inlet port into the cavity by way of a pump **116** which may be, for example, a model no. B-120-S made by Eldex Laboratories. Selected fluids are circulated into the cavity by the pump, through the cavity, and out the outlet for recirculation or disposal. The reactor may be subjected to ultrasonic radiation and/or heated to aid in agitation in some embodiments.

Above the substrate **112**, a lens **120** is provided which may be, for example, a 2" 100 mm focal length fused silica lens. For the sake of a compact system, a reflective mirror **122** may be provided for directing light from a light source **124** onto the substrate. Light source **124** may be, for example, a Xe(Hg) light source manufactured by Oriel and having model no. 66024. A second lens **126** may be provided for the purpose of projecting a mask image onto the substrate in combination with lens **112**. This form of lithography is referred to herein as projection printing. As will be apparent from this disclosure, proximity printing and the like may also be used according to some embodiments.

Light from the light source is permitted to reach only selected locations on the substrate as a result of mask **128**.

22

Mask **128** may be, for example, a glass slide having etched chrome thereon. The mask **128** in one embodiment is provided with a grid of transparent locations and opaque locations. Such masks may be manufactured by, for example, Photo Sciences, Inc. Light passes freely through the transparent regions of the mask, but is reflected from or absorbed by other regions. Therefore, only selected regions of the substrate are exposed to light.

As discussed above, light valves (LCD's) may be used as an alternative to conventional masks to selectively expose regions of the substrate. Alternatively, fiber optic faceplates such as those available from Schott Glass, Inc. may be used for the purpose of contrast enhancement of the mask or as the sole means of restricting the region to which light is applied. Such faceplates would be placed directly above or on the substrate in the reactor shown in FIG. 8A. In still further embodiments, flys-eye lenses, tapered fiber optic faceplates, or the like, may be used for contrast enhancement.

In order to provide for illumination of regions smaller than a wavelength of light, more elaborate techniques may be utilized. For example, according to one preferred embodiment, light is directed at the substrate by way of molecular microcrystals on the tip of, for example, micropipettes. Such devices are disclosed in Lieberman et al., "A Light Source Smaller Than the Optical Wavelength," *Science* (1990) 247:59-61, which is incorporated herein by reference for all purposes.

In operation, the substrate is placed on the cavity and sealed thereto. All operations in the process of preparing the substrate are carried out in a room lit primarily or entirely by light of a wavelength outside of the light range at which the protective group is removed. For example, in the case of NVOC, the room should be lit with a conventional dark room light which provides little or no UV light. All operations are preferably conducted at about room temperature.

A first, deprotection fluid (without a monomer) is circulated through the cavity. The solution preferably is of 5 mM sulfuric acid in dioxane solution which serves to keep exposed amino groups protonated and decreases their reactivity with photolysis by-products. Absorptive materials such as N,N-diethylamino 2,4-dinitrobenzene, for example, may be included in the deprotection fluid which serves to absorb light and prevent reflection and unwanted photolysis.

The slide is, thereafter, positioned in a light raypath from the mask such that first locations on the substrate are illuminated and, therefore, deprotected. In preferred embodiments the substrate is illuminated for between about 1 and 15 minutes with a preferred illumination time of about 10 minutes at 10-20 mW/cm² with 365 nm light. The slides are neutralized (i.e., brought to a pH of about 7) after photolysis with, for example, a solution of di-isopropylethylamine (DIEA) in methylene chloride for about 5 minutes.

The first monomer is then placed at the first locations on the substrate. After irradiation, the slide is removed, treated in bulk, and then reinstalled in the flow cell. Alternatively, a fluid containing the first monomer, preferably also protected by a protective group, is circulated through the cavity by way of pump **116**. If, for example, it is desired to attach the amino acid Y to the substrate at the first locations, the amino acid Y (bearing a protective group on its α -nitrogen), along with reagents used to render the monomer-reactive, and/or a carrier, is circulated from a storage container **118**, through the pump, through the cavity, and back to the inlet of the pump.

The monomer carrier solution is, in a preferred embodiment, formed by mixing of a first solution (referred

to herein as solution “A”) and a second solution (referred to herein as solution “B”). Table 2 provides an illustration of a mixture which may be used for solution A.

TABLE 2

Representative Monomer Carrier Solution “A”	
100 mg	NVOC amino protected amino acid
37 mg	HOBt (1-Hydroxybenzotriazole)
250 μl	DMF (Dimethylformamide)
86 μl	DIEA (Diisopropylethylamine)

The composition of solution B is illustrated in Table 3. Solutions A and B are mixed and allowed to react at room temperature for about 8 minutes, then diluted with 2 ml of DMF, and 500 μl are applied to the surface of the slide or the solution is circulated through the reactor system and allowed to react for about 2 hours at room temperature. The slide is then washed with DMF, methylene chloride and ethanol.

TABLE 3

Representative Monomer Carrier Solution “B”	
250 μl	DMF
111 mg	BOP (Benzotriazolyl-n-oxy-tris (dimethylamino) phosphoniumhexafluorophosphate)

As the solution containing the monomer to be attached is circulated through the cavity, the amino acid or other monomer will react at its carboxy terminus with amino groups on the regions of the substrate which have been deprotected. Of course, while the invention is illustrated by way of circulation of the monomer through the cavity, the invention could be practiced by way of removing the slide from the reactor and submersing it in an appropriate monomer solution.

After addition of the first monomer, the solution containing the first amino acid is then purged from the system. After circulation of a sufficient amount of the DMF/methylene chloride such that removal of the amino acid can be assured (e.g., about 50x times the volume of the cavity and carrier lines), the mask or substrate is repositioned, or a new mask is utilized such that second regions on the substrate will be exposed to light and the light 124 is engaged for a second exposure. This will deprotect second regions on the substrate and the process is repeated until the desired polymer sequences have been synthesized.

The entire derivatized substrate is then exposed to a receptor of interest, preferably labeled with, for example, a fluorescent marker, by circulation of a solution or suspension of the receptor through the cavity or by contacting the surface of the slide in bulk. The receptor will preferentially bind to certain regions of the substrate which contain complementary sequences.

Antibodies are typically suspended in what is commonly referred to as “supercocktail,” which may be, for example, a solution of about 1% BSA (bovine serum albumin), 0.5% Tween in PBS (phosphate buffered saline) buffer. The antibodies are diluted into the supercocktail buffer to a final concentration of, for example, about 0.1 to 4 μg/ml.

FIG. 12B illustrates an alternative preferred embodiment of the reactor shown in FIG. 8A. According to this embodiment, the mask 128 is placed directly in contact with the substrate. Preferably, the etched portion of the mask is placed face down so as to reduce the effects of light dispersion. According to this embodiment, the imaging lenses 120 and 126 are not necessary because the mask is brought into close proximity with the substrate.

For purposes of increasing the signal-to-noise ratio of the technique, some embodiments of the invention provide for exposure of the substrate to a first labeled or unlabeled receptor followed by exposure of a labeled, second receptor (e.g., an antibody) which binds at multiple sites on the first receptor. If, for example, the first receptor is an antibody derived from a first species of an animal, the second receptor is an antibody derived from a second species directed to epitopes associated with the first species. In the case of a mouse antibody, for example, fluorescently labeled goat antibody or antiserum which is antimouse may be used to bind at multiple sites on the mouse antibody, providing several times the fluorescence compared to the attachment of a single mouse antibody at each binding site. This process may be repeated again with additional antibodies (e.g., goat-mouse-goat, etc.) for further signal amplification.

In preferred embodiments an ordered sequence of masks is utilized. In some embodiments it is possible to use as few as a single mask to synthesize all of the possible polymers of a given monomer set.

If, for example, it is desired to synthesize all 16 dinucleotides from four bases, a 1 cm square synthesis region is divided conceptually into 16 boxes, each 0.25 cm wide. Denote the four monomer units by A, B, C, and D. The first reactions are carried out in four vertical columns, each 0.25 cm wide. The first mask exposes the left-most column of boxes, where A is coupled. The second mask exposes the next column, where B is coupled; followed by a third mask, for the C column; and a final mask that exposes the right-most column, for D. The first, second, third, and fourth masks may be a single mask translated to different locations.

The process is repeated in the horizontal direction for the second unit of the dimer. This time, the masks allow exposure of horizontal rows, again 0.25 cm wide. A, B, C, and D are sequentially coupled using masks that expose horizontal fourths of the reaction area. The resulting substrate contains all 16 dinucleotides of four bases.

The eight masks used to synthesize the dinucleotide are related to one another by translation or rotation. In fact, one mask can be used in all eight steps if it is suitably rotated and translated. For example, in the example above, a mask with a single transparent region could be sequentially used to expose each of the vertical columns, translated 9°, and then sequentially used to allow exposure of the horizontal rows.

Tables 4 and 5 provide a simple computer program in Quick Basic for planning a masking program and a sample output, respectively, for the synthesis of a polymer chain of three monomers (“residues”) having three different monomers in the first level, four different monomers in the second level, and five different monomers in the third level in a striped pattern. The output of the program is the number of cells, the number of “stripes” (light regions) on each mask, and the amount of translation required for each exposure of the mask.

TABLE 4

Mask Strategy Program	
DEFINT A-Z	
DIM b(20), w(20), l(500)	
F\$ = “LPT1:”	
OPEN F\$ FOR OUTPUT AS #1	
jmax = 3 Number of residues	
b(1) = 3: b(2) = 4: b(3) = 5 *Number of building blocks for res 1,2,3	
g = 1: lmax(1) = 1	
FOR j = 1 TO jmax: g = g * b(j): NEXT j	
w(0) = 0: w(1) = g / b(1)	

TABLE 4-continued

Mask Strategy Program	
<pre>PRINT #1, "MASK2.BAS", DATE\$, TIME\$: PRINT #1, PRINT #1, USING "Number of residues=##"; jmax FOR j = 1 TO jmax PRINT #1, USING " Residue ## ## building blocks"; j; b(j) NEXT j PRINT #1, " PRINT #1, USING "Number of cells=####"; g: PRINT #1, FOR j = 2 TO jmax lmax(j) = lmax(j - 1) * b(j - 1) w(j) = w(j - 1) / b(j) NEXT j FOR j = 1 TO jmax PRINT #1, USING "Mask for residue ##"; j: PRINT #1, PRINT #1, USING " Number of stripes=####"; lmax(j) PRINT #1, USING " Width of each stripe=####"; w(j) FOR l = 1 TO lmax(j) a = 1 + (l - 1) * w(j - 1) ac = a + w(j) - 1 PRINT #1, USING " Stripe ## begins at location ### and ends at ###"; l; a; ac NEXT l PRINT #1, PRINT #1, USING " For each of ## building blocks, translate mask by ## cell(s)"; b(j); w(j), PRINT #1, : PRINT #1, : PRINT #1, NEXT j</pre>	

TABLE 5

Masking Strategy Output	
Number of residues= 3	
Residue 1	3 building blocks
Residue 2	4 building blocks
Residue 3	5 building blocks
Number of cells= 60	
Mask for residue 1	
Number of stripes= 1	
Width of each stripe= 20	
Stripe 1 begins at location 1 and ends at 20	
For each of 3 building blocks, translate mask by 20 cell(s)	
Mask for residue 2	
Number of stripes= 3	
Width of each stripe= 5	
Stripe 1 begins at location 1 and ends at 5	
Stripe 2 begins at location 21 and ends at 25	
Stripe 3 begins at location 41 and ends at 45	
For each of 4 building blocks, translate mask by 5 cell(s)	
Mask for residue 3	
Number of stripes= 12	
Width of each stripe= 1	
Stripe 1 begins at location 1 and ends at 1	
Stripe 2 begins at location 6 and ends at 6	
Stripe 3 begins at location 11 and ends at 11	
Stripe 4 begins at location 16 and ends at 16	
Stripe 5 begins at location 21 and ends at 21	
Stripe 6 begins at location 26 and ends at 26	
Stripe 7 begins at location 31 and ends at 31	
Stripe 8 begins at location 36 and ends at 36	
Stripe 9 begins at location 41 and ends at 41	
Stripe 10 begins at location 46 and ends at 46	
Stripe 11 begins at location 51 and ends at 51	
Stripe 12 begins at location 56 and ends at 56	
For each of 5 building blocks, translate mask by 1 cell(s)	

V. Details of One Embodiment of a Fluorescent Detection Device

FIG. 13 illustrates a fluorescent detection device for detecting fluorescently labeled receptors on a substrate. A substrate 112 is placed on an x/y translation table 202. In a preferred embodiment the x/y translation table is a model no. PM500-A1 manufactured by Newport Corporation. The x/y translation table is connected to and controlled by an appropriately programmed digital computer 204 which may be,

for example, an appropriately programmed IBM PC/AT or AT compatible computer. Of course, other computer systems, special purpose hardware, or the like could readily be substituted for the AT computer used herein for illustration. Computer software for the translation and data collection functions described herein can be provided based on commercially available software including, for example, "Lab Windows" licensed by National Instruments, which is incorporated herein by reference for all purposes.

The substrate and x/y translation table are placed under a microscope 206 which includes one or more objectives 208. Light (about 488 nm) from a laser 210, which in some embodiments is a model no. 2020-05 argon ion laser manufactured by Spectraphysics, is directed at the substrate by a dichroic mirror 207 which passes greater than about 520 nm light but reflects 488 nm light. Dichroic mirror 207 may be, for example, a model no. FT510 manufactured by Carl Zeiss. Light reflected from the mirror then enters the microscope 206 which may be, for example, a model no. Axioscop 20 manufactured by Carl Zeiss. Fluorescein-marked materials on the substrate will fluoresce >488 nm light, and the fluoresced light will be collected by the microscope and passed through the mirror. The fluorescent light from the substrate is then directed through a wavelength filter 209 and, thereafter through an aperture plate 211. Wavelength filter 209 may be, for example, a model no. OG530 manufactured by Melles Griot and aperture plate 211 may be, for example, a model no. 477352/477380 manufactured by Carl Zeiss.

The fluoresced light then enters a photomultiplier tube 212 which in some embodiments is a model no. R943-02 manufactured by Hamamatsu, the signal is amplified in preamplifier 214 and photons are counted by photon counter 216. The number of photons is recorded as a function of the location in the computer 204. Pre-Amp 214 may be, for example, a model no. SR440 manufactured by Stanford Research Systems and photon counter 216 may be a model no. SR400 manufactured by Stanford Research Systems. The substrate is then moved to a subsequent location and the process is repeated. In preferred embodiments the data are acquired every 1 to 100 μ m with a data collection diameter of about 0.8 to 10 μ m preferred. In embodiments with sufficiently high fluorescence, a CCD detector with broad-field illumination is utilized.

By counting the number of photons generated in a given area in response to the laser, it is possible to determine where fluorescent marked molecules are located on the substrate. Consequently, for a slide which has a matrix of polypeptides, for example, synthesized on the surface thereof, it is possible to determine which of the polypeptides is complementary to a fluorescently marked receptor.

According to preferred embodiments, the intensity and duration of the light applied to the substrate is controlled by varying the laser power and scan stage rate for improved signal-to-noise ratio by maximizing fluorescence emission and minimizing background noise.

While the detection apparatus has been illustrated primarily herein with regard to the detection of marked receptors, the invention will find application in other areas. For example, the detection apparatus disclosed herein could be used in the fields of catalysis, DNA or protein gel scanning, and the like.

In the use of autoradiography, the marker is a radioactive label, such as ³²P. The marker is exposed to a film, which is developed, then read out on a scanner. An exposure time of about 1 hour would be required for autoradiography in one embodiment. Fluorescence detection using, for example, a

US 6,403,320 B1

27

fluorophore such as fluorescein attached to the receptor will usually require shorter exposure e.g., 1 second or less.

VI. Determination of Relative Binding Strength of Receptors

The signal-to-noise ratio of the present invention is sufficiently high that not only can the presence or absence of a receptor on a ligand be detected, but also the relative binding affinity of receptors to a variety of sequences can be determined.

In practice it is found that a receptor will bind to several peptide sequences in an array, but will bind much more strongly to some sequences than others. Strong binding affinity will be evidenced herein by a strong fluorescent or radiographic signal since many receptor molecules will bind in a region of a strongly bound ligand. Conversely, a weak binding affinity will be evidenced by a weak fluorescent or radiographic signal due to the relatively small number of receptor molecules which bind in a particular region of a substrate having a ligand with a weak binding affinity for the receptor. Consequently, it becomes possible to determine relative binding avidity (or affinity in the case of univalent interactions) of a ligand herein by way of the intensity of a fluorescent or radiographic signal in a region containing that ligand.

Semiquantitative data on affinities might also be obtained by varying washing conditions and concentrations of the receptor. This would be done by comparison to known ligand receptor pairs, for example.

VII. Examples

The following examples are provided to illustrate the efficacy of the inventions herein. All operations were conducted at about ambient temperatures and pressures unless indicated to the contrary.

A. Slide Preparation

Before attachment of reactive groups it is preferred to clean the substrate which is, in a preferred embodiment a glass substrate such as a microscope slide or cover slip. According to one embodiment the slide is soaked in an alkaline bath consisting of, for example, 1 liter of 95% ethanol with 120 ml of water and 120 grams of sodium hydroxide for 12 hours. The slides are then washed under running water and allowed to air dry, and rinsed once with a solution of 95% ethanol.

The slides are then aminated with, for example, aminopropyltriethoxysilane for the purpose of attaching amino groups to the glass surface on linker molecules, although any omega functionalized silane could also be used for this purpose. In one embodiment 0.1% aminopropyltriethoxysilane is utilized, although solutions with concentrations from 10⁻⁷% to 10% may be used, with about 10⁻³% to 2% preferred. A 0.1% mixture is prepared by adding to 100 ml of a 95% ethanol/5% water mixture, 100 microliters (μ l) of aminopropyltriethoxysilane. The mixture is agitated at about ambient temperature on a rotary shaker for about 5 minutes. 500 μ l of this mixture is then applied to the surface of one side of each cleaned slide. After 4 minutes, the slides are decanted of this solution and rinsed three times by dipping in, for example, 100% ethanol.

After the plates dry, they are placed in a 110–1200° C. vacuum oven for about 20 minutes, and then allowed to cure at room temperature for about 12 hours in an argon environment. The slides are then dipped into DMF (dimethylformamide) solution, followed by a thorough washing with methylene chloride.

The aminated surface of the slide is then exposed to about 500 μ l of, for example, a 30 millimolar (mM) solution of NVOC-GABA (gamma amino butyric acid) NHS

28

(N-hydroxysuccinimide) in DMF for attachment of a NVOC-GABA to each of the amino groups.

The surface is washed with, for example, DMF, methylene chloride, and ethanol.

Any unreacted aminopropyl silane on the surface—that is, those amino groups which have not had the NVOC-GABA attached—are now capped with acetyl groups (to prevent further reaction) by exposure to a 1:3 mixture of acetic anhydride in pyridine for 1 hour. Other materials which may perform this residual capping function include trifluoroacetic anhydride, formicacetic anhydride, or other reactive acylating agents. Finally, the slides are washed again with DMF, methylene chloride, and ethanol.

B. Synthesis of Eight Trimers of “A” and “B”

FIG. 14 illustrates a possible synthesis of the eight trimers of the two-monomer set: gly, phe (represented by “A” and “B,” respectively). A glass slide bearing silane groups terminating in 6-nitro-veratryloxycarboxamide (NVOC-UH) residues is prepared as a substrate. Active esters (pentafluorophenyl, OBt, etc.) of gly and phe protected at the amino group with NVOC are prepared as reagents. While not pertinent to this example, if side chain protecting groups are required for the monomer set, these must not be photo-reactive at the wavelength of light used to protect the primary chain.

For a monomer set of size n, n×1 cycles are required to synthesize all possible sequences of length l. A cycle consists of:

1. Irradiation through an appropriate mask to expose the amino groups at the sites where the next residue is to be added, with appropriate washes to remove the by-products of the deprotection.

2. Addition of a single activated and protected (with the same photochemically-removable group) monomer, which will react only at the sites addressed in step 1, with appropriate washes to remove the excess reagent from the surface.

The above cycle is repeated for each member of the monomer set until each location on the surface has been extended by one residue in one embodiment. In other embodiments, several residues are sequentially added at one location before moving on to the next location. Cycle times will generally be limited by the coupling reaction rate, now as short as 20 min in automated peptide synthesizers. This step is optionally followed by addition of a protecting group to stabilize the array for later testing. For some types of polymers (e.g., peptides), a final deprotection of the entire surface (removal of photoprotective side chain groups) may be required.

More particularly, as shown in FIG. 14A, the glass 20 is provided with regions 22, 24, 26, 28, 30, 32, 34, and 36. Regions 30, 32, 34, and 36 are masked, as shown in FIG. 14B and the glass is irradiated and exposed to a reagent containing “A” (e.g., gly), with the resulting structure shown in FIG. 14C. Thereafter, regions 22, 24, 26, and 28 are masked, the glass is irradiated (as shown in FIG. 14D) and exposed to a reagent containing “B” (e.g., phe), with the resulting structure shown in FIG. 14E. The process proceeds, consecutively masking and exposing the sections as shown until the structure shown in FIG. 14M is obtained. The glass is irradiated and the terminal groups are, optionally, capped by acetylation. As shown, all possible trimers of gly/phe are obtained.

In this example, no side chain protective group removal is necessary. If it is desired, side chain deprotection may be accomplished by treatment with ethanedithiol and trifluoroacetic acid.

In general, the number of steps needed to obtain a particular polymer chain is defined by:

US 6,403,320 B1

29

$$n \times l \quad (1)$$

where:

n =the number of monomers in the basis set of monomers, and

l =the number of monomer units in a polymer chain.

Conversely, the synthesized number of sequences of length l will be:

$$n^l. \quad (2)$$

Of course, greater diversity is obtained by using masking strategies which will also include the synthesis of polymers having a length of less than l . If, in the extreme case, all polymers having a length less than or equal to l are synthesized, the number of polymers synthesized will be:

$$n^l + n^{l-1} + \dots + n^1. \quad (3)$$

The maximum number of lithographic steps needed will generally be n for each "layer" of monomers, i.e., the total number of masks (and, therefore, the number of lithographic steps) needed will be $n \times l$. The size of the transparent mask regions will vary in accordance with the area of the substrate available for synthesis and the number of sequences to be formed. In general, the size of the synthesis areas will be:

$$\text{size of synthesis areas} = (A)/(S)$$

where:

A is the total area available for synthesis; and

S is the number of sequences desired in the area.

It will be appreciated by those of skill in the art that the above method could readily be used to simultaneously produce thousands or millions of oligomers on a substrate using the photolithographic techniques disclosed herein. Consequently, the method results in the ability to practically test large numbers of, for example, di, tri, tetra, penta, hexa, hepta, octapeptides, dodecapeptides, or larger polypeptides (or correspondingly, polynucleotides).

The above example has illustrated the method by way of a manual example. It will of course be appreciated that automated or semi-automated methods could be used. The substrate would be mounted in a flow cell for automated addition and removal of reagents, to minimize the volume of reagents needed, and to more carefully control reaction conditions. Successive masks could be applied manually or automatically.

C. Synthesis of a Dimer of an Aminopropyl Group and a Fluorescent Group

In synthesizing the dimer of an aminopropyl group and a fluorescent group, a functionalized durapore membrane was used as a substrate. The durapore membrane was a polyvinylidene difluoride with aminopropyl groups. The aminopropyl groups were protected with the DDZ group by reaction of the carbonyl chloride with the amino groups, a reaction readily known to those of skill in the art. The surface bearing these groups was placed in a solution of THF and contacted with a mask bearing a checkerboard pattern of 1 mm opaque and transparent regions. The mask was exposed to ultraviolet light having a wavelength down to at least about 280 nm for about 5 minutes at ambient temperature, although a wide range of exposure times and temperatures may be appropriate in various embodiments of the invention. For example, in one embodiment, an exposure time of between about 1 and 5000 seconds may be used at process temperatures of between -70 and $+500^\circ \text{C}$.

30

In one preferred embodiment, exposure times of between about 1 and 500 seconds at about ambient pressure are used. In some preferred embodiments, pressure above ambient is used to prevent evaporation.

The surface of the membrane was then washed for about 1 hour with a fluorescent label which included an active ester bound to a chelate of a lanthanide. Wash times will vary over a wide range of values from about a few minutes to a few hours. These materials fluoresce in the red and the green visible region. After the reaction with the active ester in the fluorophore was complete, the locations in which the fluorophore was bound could be visualized by exposing them to ultraviolet light and observing the red and the green fluorescence. It was observed that the derivatized regions of the substrate closely corresponded to the original pattern of the mask.

D. Demonstration of Signal Capability

Signal detection capability was demonstrated using a low-level standard fluorescent bead kit manufactured by Flow Cytometry Standards and having model no. 824. This kit includes $5.8 \mu\text{m}$ diameter beads, each impregnated with a known number of fluorescein molecules.

One of the beads was placed in the illumination field on the scan stage as shown in FIG. 9 in a field of a laser spot which was initially shuttered. After being positioned in the illumination field, the photon detection equipment was turned on. The laser beam was unblocked and it interacted with the particle bead, which then fluoresced. Fluorescence curves of beads impregnated with 7,000; 13,000; and 29,000 fluorescein molecules, are shown in FIGS. 11A, 11B, and 11C respectively. On each curve, traces for beads without fluorescein molecules are also shown. These experiments were performed with 488 nm excitation, with 100 μW of laser power. The light was focused through a 40 power 0.75 NA objective.

The fluorescence intensity in all cases started off at a high value and then decreased exponentially. The fall-off in intensity is due to photobleaching of the fluorescein molecules. The traces of beads without fluorescein molecules are used for background subtraction. The difference in the initial exponential decay between labeled and nonlabeled beads is integrated to give the total number of photon counts, and this number is related to the number of molecules per bead. Therefore, it is possible to deduce the number of photons per fluorescein molecule that can be detected. For the curves illustrated in FIG. 11, this calculation indicates the radiation of about 40 to 50 photons per fluorescein molecule are detected.

E. Determination of the Number of Molecules Per Unit Area

Aminopropylated glass microscope slides prepared according to the methods discussed above were utilized in order to establish the density of labeling of the slides. The free amino termini of the slides were reacted with FITC (fluorescein isothiocyanate) which forms a covalent linkage with the amino group. The slide is then scanned to count the number of fluorescent photons generated in a region which, using the estimated 40–50 photons per fluorescein molecule, enables the calculation of the number of molecules which are on the surface per unit area.

A slide with aminopropyl silane on its surface was immersed in a 1 mM solution of FITC in DMF for 1 hour at about ambient temperature. After reaction, the slide was washed twice with DMF and then washed with ethanol, water, and then ethanol again. It was then dried and stored in the dark until it was ready to be examined.

Through the use of curves similar to those shown in FIG. 15, and by integrating the fluorescent counts under the

US 6,403,320 B1

31

exponentially decaying signal, the number of free amino groups on the surface after derivitization was determined. It was determined that slides with labeling densities of 1 fluoroscein per $10^3 \times 10^3$ to $\sim 2 \times 2$ nm could be reproducibly made as the concentration of aminopropyltriethoxysilane

F. Removal of NVOC and Attachment of a Fluorescent Marker

NVOC-GABA groups were attached as described above. The entire surface of one slide was exposed to light so as to expose a free amino group at the end of the gamma amino butyric acid. This slide, and a duplicate which was not exposed, were then exposed to fluorescein isothiocyanate (FITC).

FIG. 16A illustrates the slide which was not exposed to light, but which was exposed to FITC. The units of the x axis are time and the units of the y axis are counts. The trace contains a certain amount of background fluorescence. The duplicate slide was exposed to 350 nm broadband illumination for about 1 minute (12 mW/cm^2 , ~ 350 nm illumination), washed and reacted with FITC. The fluorescence curves for this slide are shown in FIG. 16B. A large increase in the level of fluorescence is observed, which indicates photolysis has exposed a number of amino groups on the surface of the slides for attachment of a fluorescent marker.

G. Use of a Mask in Removal of NVOC

The next experiment was performed with a 0.1% aminopropylated slide. Light from a Hg—Xe arc lamp was imaged onto the substrate through a laser-ablated chrome-on-glass mask in direct contact with the substrate.

This slide was illuminated for approximately 5 minutes, with 12 mW of 350 nm broadband light and then reacted with the 1 mM FITC solution. It was put on the laser detection scanning stage and a graph was plotted as a two-dimensional representation of position color-coded for fluorescence intensity. The fluorescence intensity (in counts) as a function of location is given on the color scale to the right of FIG. 17A for a mask having $100 \times 100 \mu\text{m}$ squares.

The experiment was repeated a number of times through various masks. The fluorescence pattern for a $50 \mu\text{m}$ mask is illustrated in FIG. 17B, for a $20 \mu\text{m}$ mask in FIG. 17C, and for a $10 \mu\text{m}$ mask in FIG. 17D. The mask pattern is distinct down to at least about $10 \mu\text{m}$ squares using this lithographic technique.

H. Attachment of YGGFL and Subsequent Exposure to Herz Antibody and Goat Antimouse

In order to establish that receptors to a particular polypeptide sequence would bind to a surface-bound peptide and be detected, Leu enkephalin was coupled to the surface and recognized by an antibody. A slide was derivatized with 0.1% amino propyl-triethoxysilane and protected with NVOC. A $500 \mu\text{m}$ checkerboard mask was used to expose the slide in a flow cell using backside contact printing. The Leu enkephalin sequence (H_2N -tyrosine, glycine, glycine, phenylalanine, leucine- CO_2H , otherwise referred to herein as YGGFL) was attached via its carboxy end to the exposed amino groups on the surface of the slide. The peptide was added in DMF solution with the BOP/HOBT/DIEA coupling reagents and recirculated through the flow cell for 2 hours at room temperature.

A first antibody, known as the Herz antibody, was applied to the surface of the slide for 45 minutes at $2 \mu\text{g/ml}$ in a supercocktail (containing 1% BSA and 1% ovalbumin also in this case). A second antibody, goat anti-mouse fluorescein conjugate, was then added at $2 \mu\text{g/ml}$ in the supercocktail buffer, and allowed to incubate for 2 hours.

32

The results of this experiment are provided in FIG. 18. Again, this figure illustrates fluorescence intensity as a function of position. The fluorescence scale is shown on the right, according to the color coding. This image was taken at $10 \mu\text{m}$ steps. This figure indicates that not only can deprotection be carried out in a well defined pattern, but also that (1) the method provides for successful coupling of peptides to the surface of the substrate, (2) the surface of a bound peptide is available for binding with an antibody, and (3) that the detection apparatus capabilities are sufficient to detect binding of a receptor.

I. Monomer-by-Monomer Formation of YGGFL and Subsequent Exposure to Labeled Antibody

Monomer-by-monomer synthesis of YGGFL and GGFL in alternate squares was performed on a slide in a checkerboard pattern and the resulting slide was exposed to the Herz antibody. This experiment and the results thereof are illustrated in FIGS. 19A, 19B, 19C, and 19D.

In FIG. 19A, a slide is shown which is derivatized with the aminopropyl group, protected in this case with t-BOC (t-butoxycarbonyl). The slide was treated with TFA to remove the t-BOC protecting group. E-aminocaproic acid, which was t-BOC protected at its amino group, was then coupled onto the aminopropyl groups. The aminocaproic acid serves as a spacer between the aminopropyl group and the peptide to be synthesized. The amino end of the spacer was deprotected and coupled to NVOC-leucine. The entire slide was then illuminated with 12 mW of 325 nm broadband illumination. The slide was then coupled with NVOC-phenylalanine and washed. The entire slide was again illuminated, then coupled to NVOC-glycine and washed. The slide was again illuminated and coupled to NVOC-glycine to form the sequence shown in the last portion of FIG. 19A.

As shown in FIG. 19B, alternating regions of the slide were then illuminated using a projection print using a $500 \times 500 \mu\text{m}$ checkerboard mask; thus, the amino group of glycine was exposed only in the lighted areas. When the next coupling chemistry step was carried out, NVOC-tyrosine was added, and it coupled only at those spots which had received illumination. The entire slide was then illuminated to remove all the NVOC groups, leaving a checkerboard of YGGFL in the lighted areas and in the other areas, GGFL. The Herz antibody (which recognizes the YGGFL, but not GGFL) was then added, followed by goat anti-mouse fluorescein conjugate.

The resulting fluorescence scan is shown in FIG. 19C, and the color coding for the fluorescence intensity is again given on the right. Dark areas contain the tetrapeptide GGFL, which is not recognized by the Herz antibody (and thus there is no binding of the goat anti-mouse antibody with fluorescein conjugate), and in the red areas YGGFL is present. The YGGFL pentapeptide is recognized by the Herz antibody and, therefore, there is antibody in the lighted regions for the fluorescein-conjugated goat anti-mouse to recognize.

Similar patterns are shown for a $50 \mu\text{m}$ mask used in direct contact ("proximity print") with the substrate in FIG. 19D. Note that the pattern is more distinct and the corners of the checkerboard pattern are touching when the mask is placed in direct contact with the substrate (which reflects the increase in resolution using this technique).

J. Monomer-by-Monomer Synthesis of YGGFL and PGGFL

A synthesis using a $50 \mu\text{m}$ checkerboard mask similar to that shown in FIG. 19 was conducted. However, P was added to the GGFL sites on the substrate through an additional coupling step. P was added by exposing protected GGFL to light through a mask, and subsequent exposure to P in the

manner set forth above. Therefore, half of the regions on the substrate contained YGGFL and the remaining half contained PGGFL.

The fluorescence plot for this experiment is provided in FIG. 20. As shown, the regions are again readily discernable. This experiment demonstrates that antibodies are able to recognize a specific sequence and that the recognition is not length-dependent.

K. Monomer-by-Monomer Synthesis of YGGFL and YPGGFL

In order to further demonstrate the operability of the invention, a 50 μ m checkerboard pattern of alternating YGGFL and YPGGFL was synthesized on a substrate using techniques like those set forth above. The resulting fluorescence plot is provided in FIG. 21. Again, it is seen that the antibody is clearly able to recognize the YGGFL sequence and does not bind significantly at the YPGGFL regions.

L. Synthesis of an Array or Sixteen Different Amino Acid Sequences and Estimation of Relative Binding Affinity to Herz Antibody

Using techniques similar to those set forth above, an array of 16 different amino acid sequences (replicated four times) was synthesized on each of two glass substrates. The sequences were synthesized by attaching the sequence NVOC-GFL across the entire surface of the slides. Using a series of masks, two layers of amino acids were then selectively applied to the substrate. Each region had dimensions of 0.25 cm \times 0.0625 cm. The first slide contained amino acid sequences containing only L amino acids while the second slide contained selected D amino acids. FIGS. 18A and 18B illustrate a map of the various regions on the first and second slides, respectively. The patterns shown in FIGS. 22A and 22B were duplicated four times on each slide. The slides were then exposed to the Herz antibody and fluorescein-labeled goat anti-mouse.

FIG. 23 is a fluorescence plot of the first slide, which contained only L amino acids. Red indicates strong binding (149,000 counts or more) while black indicates little or no binding of the Herz antibody (20,000 counts or less). The bottom right-hand portion of the slide appears "cut off" because the slide was broken during processing. The sequence YGGFL is clearly most strongly recognized. The sequences YAGFL and YSGFL also exhibit strong recognition of the antibody. By contrast, most of the remaining sequences show little or no binding. The four duplicate portions of the slide are extremely consistent in the amount of binding shown therein.

FIG. 24 is a fluorescence plot of the second slide. Again, strongest binding is exhibited by the YGGFL sequence. Significant binding is also detected to YaGFL, YsGFL, and YpGFL. The remaining sequences show less binding with the antibody. Note the low binding efficiency of the sequence yGGFL.

Table 6 lists the various sequences tested in order of relative fluorescence, which provides information regarding relative binding affinity.

TABLE 6

Apparent Binding to Herz Ab	
L-a.a. Set	D-a.a. Set
YGGFL	YGGFL
YAGFL	YaGFL
YSGFL	YsGFL
LGGFL	YpGFL
FGGFL	fGGFL

TABLE 6-continued

Apparent Binding to Herz Ab	
L-a.a. Set	D-a.a. Set
YPGFL	yGGFL
LAGFL	faGFL
FAGFL	wGGFL
WGGFL	yaGFL
	fpGFL
	waGFL

VIII. Illustrative Alternative Embodiment

According to an alternative embodiment of the invention, the methods provide for attaching to the surface a caged binding member which in its caged form has a relatively low affinity for other potentially binding species, such as receptors and specific binding substances. Such techniques are more fully described in copending application Ser. No. 404,920, filed Sep. 8, 1989, and incorporated herein by reference for all purposes.

According to this alternative embodiment, the invention provides methods for forming predefined regions on a surface of a solid support, wherein the predefined regions are capable of immobilizing receptors. The methods make use of caged binding members attached to the surface to enable selective activation of the predefined regions. The caged binding members are liberated to act as binding members ultimately capable of binding receptors upon selective activation of the predefined regions. The activated binding members are then used to immobilize specific molecules such as receptors on the predefined region of the surface. The above procedure is repeated at the same or different sites on the surface so as to provide a surface prepared with a plurality of regions on the surface containing, for example, the same or different receptors. When receptors immobilized in this way have a differential affinity for one or more ligands, screenings and assays for the ligands can be conducted in the regions of the surface containing the receptors.

The alternative embodiment may make use of novel caged binding members attached to the substrate. Caged (unactivated) members have a relatively low affinity for receptors of substances that specifically bind to uncaged binding members when compared with the corresponding affinities of activated binding members. Thus, the binding members are protected from reaction until a suitable source of energy is applied to the regions of the surface desired to be activated. Upon application of a suitable energy source, the caging groups labilize, thereby presenting the activated binding member. A typical energy source will be light.

Once the binding members on the surface are activated they may be attached to a receptor. The receptor chosen may be a monoclonal antibody, a nucleic acid sequence, a drug receptor, etc. The receptor will usually, though not always, be prepared so as to permit attaching it, directly or indirectly, to a binding member. For example, a specific binding substance having a strong binding affinity for the binding member and a strong affinity for the receptor or a conjugate of the receptor may be used to act as a bridge between binding members and receptors if desired. The method uses a receptor prepared such that the receptor retains its activity toward a particular ligand.

Preferably, the caged binding member attached to the solid substrate will be a photoactivatable biotin complex, i.e., a biotin molecule that has been chemically modified with photoactivatable protecting groups so that it has a significantly reduced binding affinity for avidin or avidin

US 6,403,320 B1

35

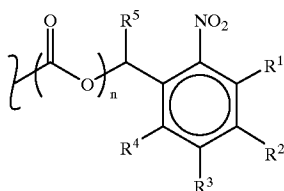
analogs than does natural biotin. In a preferred embodiment, the protecting groups localized in a predefined region of the surface will be removed upon application of a suitable source of radiation to give binding members, that are biotin or a functionally analogous compound having substantially the same binding affinity for avidin or avidin analogs as does biotin.

In another preferred embodiment, avidin or an avidin analog is incubated with activated binding members on the surface until the avidin binds strongly to the binding members. The avidin so immobilized on predefined regions of the surface can then be incubated with a desired receptor or conjugate of a desired receptor. The receptor will preferably be biotinylated, e.g., a biotinylated antibody, when avidin is immobilized on the predefined regions of the surface. Alternatively, a preferred embodiment will present an avidin/biotinylated receptor complex, which has been previously prepared, to activated binding members on the surface.

IX. Conclusion

The present inventions provide greatly improved methods and apparatus for synthesis of polymers on substrates. It is to be understood that the above description is intended to be illustrative and not restrictive. Many embodiments will be apparent to those of skill in the art upon reviewing the above description. By way of example, the invention has been described primarily with reference to the use of photoremovable protective groups, but it will be readily recognized by those of skill in the art that sources of radiation other than light could also be used. For example, in some embodiments it may be desirable to use protective groups which are sensitive to electron beam irradiation, x-ray irradiation, in combination with electron beam lithograph, or x-ray lithography techniques. Alternatively, the group could be removed by exposure to an electric current.

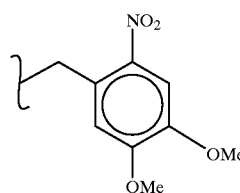
A preferred class of photoremovable protecting groups has the general formula:



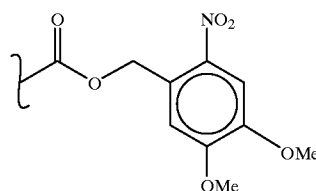
where R^1 , R^2 , R^3 , and R^4 independently are a hydrogen atom, a lower alkyl, aryl, benzyl, halogen, hydroxyl, alkoxy, thiol, thioether, amino, nitro, carboxyl, formate, formamido or phosphido group, or adjacent substituents (i.e., R^1 - R^2 , R^2 - R^3 , R^3 - R^4) are substituted oxygen groups that together form a cyclic acetal or ketal; R^5 is a hydrogen atom, a alkoxy, alkyl, hydrogen, halo, aryl, or alkenyl group, and $n=0$ or 1.

A preferred protecting group, 6-nitroveratryl (NV), which is used for protecting the carboxyl terminus of an amino acid or the hydroxyl group of a nucleotide, for example, is formed when R^2 and R^3 are each a methoxy group, R^1 , R^4 and R^5 are each a hydrogen atom, and $n=0$:

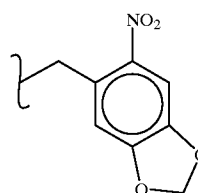
36



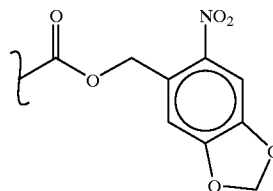
A preferred protecting group, 6-nitroveratryloxycarbonyl (NVOC), which is used to protect the amino terminus of an amino acid, for example, is formed when R^2 and R^3 are each a methoxy group, R^1 , R^4 and R^5 are each a hydrogen atom, and $n=1$:



Another preferred protecting group, 6-nitropiperonyl (NP), which is used for protecting the carboxyl terminus of an amino acid or the hydroxyl group of a nucleotide, for example, is formed when R^2 and R^3 together form a methylene acetal, R^1 , R^4 and R^5 are each a hydrogen atom, and $n=0$:



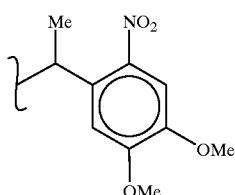
Another preferred protecting group, 6-nitropiperonyloxycarbonyl (NPOC), which is used to protect the amino terminus of an amino acid, for example, is formed when R^2 and R^3 together form a methylene acetal, R^1 , R^4 and R^5 are each a hydrogen atom, and $n=1$:



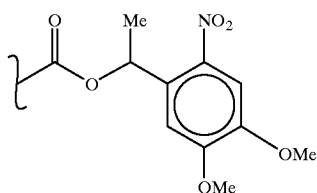
A most preferred protecting group, methyl-6-nitroveratryl (MeNV), which is used for protecting the carboxyl terminus of an amino acid or the hydroxyl group of a nucleotide, for example, is formed when R^2 and R^3 are each a methoxy group, R^1 and R^4 are each a hydrogen atom, R^5 is a methyl group, and $n=0$:

US 6,403,320 B1

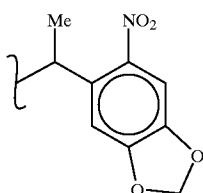
37



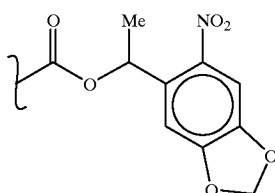
Another most preferred protecting group, methyl-6-nitroveratryloxycarbonyl (MeNVOC), which is used to protect the amino terminus of an amino acid, for example, is formed when R^2 and R^3 are each a methoxy group, R^1 and R^4 are each a hydrogen atom, R^5 is a methyl group, and $n=1$:



Another most preferred protecting group, methyl-6-nitropiperonyl (MeNP), which is used for protecting the carboxyl terminus of an amino acid or the hydroxyl group of a nucleotide, for example, is formed when R^2 and R^3 together form a methylene acetal, R^1 and R^4 are each a hydrogen atom, R^5 is a methyl group, and $n=0$:



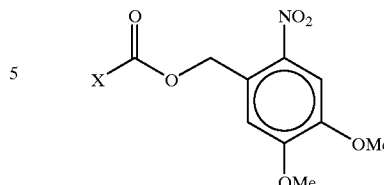
Another most preferred protecting group, methyl-6-nitropiperonyloxycarbonyl (MeNPOC), which is used to protect the amino terminus of an amino acid, for example, is formed when R^2 and R^3 together form a methylene acetal, R^1 and R^4 are each a hydrogen atom, R^5 is a methyl group, and $n=1$:



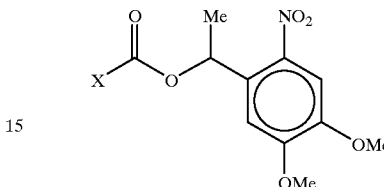
A protected amino acid having a photoactivatable oxycarbonyl protecting group, such as NVOC or NPOC or their corresponding methyl derivatives, MeNVOC or MeNPOC, respectively, on the amino terminus is formed by acylating the amine of the amino acid with an activated oxycarbonyl ester of the protecting group. Examples of activated oxycarbonyl esters of NVOC and MeNVOC have the general formula:

38

NVOC-X



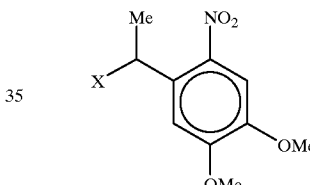
MeNVOC-X



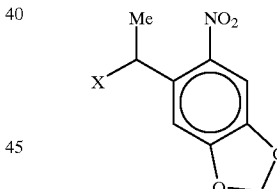
where X is halogen, mixed anhydride, phenoxy, p-nitrophenoxy, N-hydroxysuccinimide, and the like.

A protected amino acid or nucleotide having a photoactivatable protecting group, such as NV or NP or their corresponding methyl derivatives, MeNV or MeNP, respectively, on the carboxy terminus of the amino acid or 5'-hydroxy terminus of the nucleotide, is formed by acylating the carboxy terminus or 5'-OH with an activated benzyl derivative of the protecting group. Examples of activated benzyl derivatives of MeNV and MeNP have the general formula:

MeNV-X



MeNP-X



where X is halogen, hydroxyl, tosyl, mesyl, trifluoromethyl, diazo, azido, and the like.

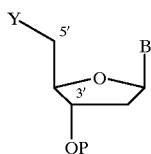
Another method for generating protected monomers is to react the benzylic alcohol derivative of the protecting group with an activated ester of the monomer. For example, to protect the carboxyl terminus of an amino acid, an activated ester of the amino acid is reacted with the alcohol derivative of the protecting group, such as 6-nitroveratrol (NVOH). Examples of activated esters suitable for such uses include halo-formate, mixed anhydride, imidazolyl formate, acyl halide, and also includes formation of the activated ester in situ the use of common reagents such as DCC and the like. See Atherton et al. for other examples of activated esters.

A further method for generating protected monomers is to react the benzylic alcohol derivative of the protecting group with an activated carbon of the monomer. For example, to protect the 5'-hydroxyl group of a nucleic acid, a derivative having a 5'-activated carbon is reacted with the alcohol derivative of the protecting group, such as methyl-6-

US 6,403,320 B1

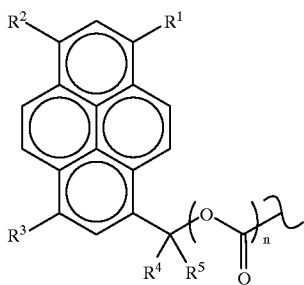
39

nitropiperonol (MePyROH). Examples of nucleotides having activating groups attached to the 5'-hydroxyl group have the general formula:



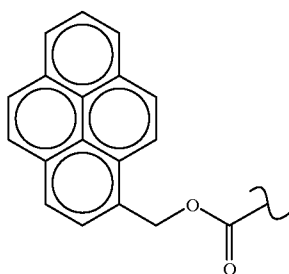
where Y is a halogen atom, a tosyl, mesyl, trifluoromethyl, azido, or diazo group, and the like.

Another class of preferred photochemical protecting groups has the formula:



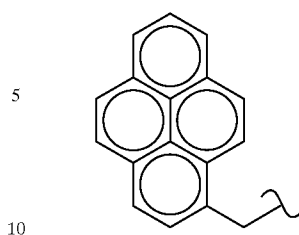
where R^1 , R^2 , and R^3 independently are a hydrogen atom, a lower alkyl, aryl, benzyl, halogen, hydroxyl, alkoxy, thiol, thioether, amino, nitro, carboxyl, formate, formamido, sulfonates, sulfido or phosphido group, R^4 and R^5 independently are a hydrogen atom, an alkoxy, alkyl, halo, aryl, hydrogen, or alkenyl group, and $n=0$ or 1.

A preferred protecting group, 1-pyrenylmethyloxycarbonyl (PyROC), which is used to protect the amino terminus of an amino acid, for example, is formed when R^1 through R^5 are each a hydrogen atom and $n=1$:

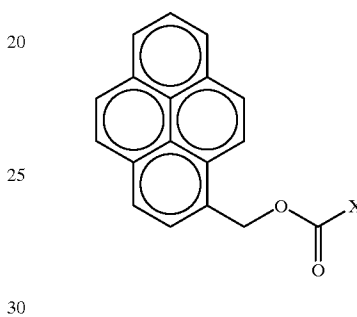


Another preferred protecting group, 1-pyrenylmethyl (PyR), which is used for protecting the carboxy terminus of an amino acid or the hydroxyl group of a nucleotide, for example, is formed when R^1 through R^5 are each a hydrogen atom and $n=0$:

40

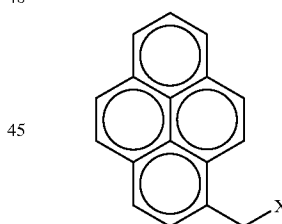


An amino acid having a pyrenylmethyloxycarbonyl protecting group on its amino terminus is formed by acylation of the free amine of amino acid with an activated oxycarbonyl ester of the pyrenyl protecting group. Examples of activated oxycarbonyl esters of PYROC have the general formula:



where X is halogen, or mixed anhydride, p-nitrophenoxy, or N-hydroxysuccinimide group, and the like.

A protected amino acid or nucleotide having a photoactivatable protecting group, such as PyR, on the carboxy terminus of the amino acid or 5'-hydroxy terminus of the nucleic acid, respectively, is formed by acylating the carboxy terminus or 5'-OH with an activated pyrenylmethyl derivative of the protecting group. Examples of activated pyrenylmethyl derivatives of PyR have the general formula:



where X is a halogen atom, a hydroxyl, diazo, or azido group, and the like.

Another method of generating protected monomers is to react the pyrenylmethyl alcohol moiety of the protecting group with an activated ester of the monomer. For example, an activated ester of an amino acid can be reacted with the alcohol derivative of the protecting group, such as pyrenylmethyl alcohol (PyROH), to form the protected derivative of the carboxy terminus of the amino acid. Examples of activated esters include halo-formate, mixed anhydride, imidazolyl formate, acyl halide, and also includes formation of the activated ester in situ and the use of common reagents such as DCC and the like.

Clearly, many photosensitive protecting groups are suitable for use in the present invention.

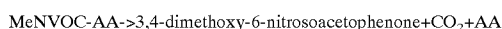
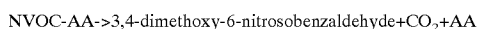
In preferred embodiments, the substrate is irradiated to remove the photoremovable protecting groups and create

US 6,403,320 B1

41

regions having free reactive moieties and side products resulting from the protecting group. The removal rate of the protecting groups depends on the wavelength and intensity of the incident radiation, as well as the physical and chemical properties of the protecting group itself. Preferred protecting groups are removed at a faster rate and with a lower intensity of radiation. For example, at a given set of conditions, MeNVOC and MeNPOC are photolytically removed from the N-terminus of a peptide chain faster than their unsubstituted parent compounds, NVOC and NPOC, respectively.

Removal of the protecting group is accomplished by irradiation to liberate the reactive group and degradation products derived from the protecting group. Not wishing to be bound by theory, it is believed that irradiation of an NVOC- and MeNVOC-protected oligomers occurs by the following reaction schemes:



where AA represents the N-terminus of the amino acid oligomer.

Along with the unprotected amino acid, other products are liberated into solution: carbon dioxide and a 2,3-dimethoxy-6-nitrosophenylcarbonyl compound, which can react with nucleophilic portions of the oligomer to form unwanted secondary reactions. In the case of an NVOC-protected amino acid, the degradation product is a nitrosobenzaldehyde, while the degradation product for the other is a nitrosophenyl ketone. For instance, it is believed that the product aldehyde from NVOC degradation reacts with free amines to form a Schiff base (imine) that affects the remaining polymer synthesis. Preferred photoremovable protecting groups react slowly or reversibly with the oligomer on the support.

Again not wishing to be bound by theory, it is believed that the product ketone from irradiation of a MeNVOC-protected oligomer reacts at a slower rate with nucleophiles on the oligomer than the product aldehyde from irradiation of the same NVOC-protected oligomer. Although not unambiguously determined, it is believed that this difference in reaction rate is due to the difference in general reactivity between aldehyde and ketones towards nucleophiles due to steric and electronic effects.

The photoremovable protecting groups of the present invention are readily removed. For example, the photolysis of N-protected L-phenylalanine in solution and having different photoremovable protecting groups was analyzed, and the results are presented in the following table:

TABLE

Photolysis of Protected L-Phe-OH

Solvent	$t_{1/2}$ in seconds			
	NBOC	NVOC	MeNVOC	MeNPOC
Dioxane	1288	110	24	19
5mM H ₂ SO ₄ /Dioxane	1575	98	33	22

The half life, $t_{1/2}$, is the time in seconds required to remove 50% of the starting amount of protecting group. NBOC is the 6-nitrobenzyloxycarbonyl group, NVOC is the 6-nitroveratryloxycarbonyl group, MeNVOC is the methyl-6-nitroveratryloxycarbonyl group, and MeNPOC is the methyl-6-nitropiperonyloxycarbonyl group. The photolysis was carried out in the indicated solvent with 362/364

42

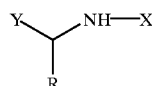
nm-wavelength irradiation having an intensity of 10 mW/cm², and the concentration of each protected phenylalanine was 0.10 mM.

The table shows that deprotection of NVOC-, MeNVOC-, and MeNPOC-protected phenylalanine proceeded faster than the deprotection of NBOC. Furthermore, it shows that the deprotection of the two derivatives that are substituted on the benzylic carbon, MeNVOC and MeNPOC, were photolyzed at the highest rates in both dioxane and acidified dioxane.

1. Use of Photoremovable Groups During Solid-Phase Synthesis of Peptides

The formation of peptides on a solid-phase support requires the stepwise attachment of an amino acid to a substrate-bound growing chain. In order to prevent unwanted polymerization of the monomeric amino acid under the reaction conditions, protection of the amino terminus of the amino acid is required. After the monomer is coupled to the end of the peptide, the N-terminal protecting group is removed, and another amino acid is coupled to the chain. This cycle of coupling and deprotecting is continued for each amino acid in the peptide sequence. See Merrifield, *J. Am. Chem. Soc.* (1963) 85:2149, and Atherton et al., "Solid Phase Peptide Synthesis" 1989, IRL Press, London, both incorporated herein by reference for all purposes. As described above, the use of a photoremovable protecting group allows removal of selected portions of the substrate surface, via patterned irradiation, during the deprotection cycle of the solid phase synthesis. This selectively allows spatial control of the synthesis—the next amino acid is coupled only to the irradiated areas.

In one embodiment, the photoremovable protecting groups of the present invention are attached to an activated ester of an amino acid at the amino terminus:



where R is the side chain of a natural or unnatural amino acid, X is a photoremovable protecting group, and Y is an activated carboxylic acid derivative. The photoremovable protecting group, X, is preferably NVOC, NPOC, PyROC, MeNVOC, MeNPOC, and the like as discussed above. The activated ester, Y, is preferably a reactive derivative having a high coupling efficiency, such as an acyl halide, mixed anhydride, N-hydroxysuccinimide ester, perfluorophenyl ester, or urethane protected acid, and the like. Other activated esters and reaction conditions are well known (See Atherton et al.).

2. Use of Photoremovable Groups During Solid-Phase Synthesis of Oligonucleotides

The formation of oligonucleotides on a solid-phase support requires the stepwise attachment of a nucleotide to a substrate-bound growing oligomer. In order to prevent unwanted polymerization of the monomeric nucleotide under the reaction conditions, protection of the 5'-hydroxyl group of the nucleotide is required. After the monomer is coupled to the end of the oligomer, the 5'-hydroxyl protecting group is removed, and another nucleotide is coupled to the chain. This cycle of coupling and deprotecting is continued for each nucleotide in the oligomer sequence. See Gait, "Oligonucleotide Synthesis: A Practical Approach" 1984, IRL Press, London, incorporated herein by reference for all purposes. As described above, the use of a photoremovable protecting group allows removal, via patterned

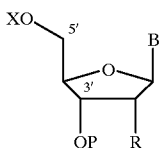
US 6,403,320 B1

43

irradiation, of selected portions of the substrate surface during the deprotection cycle of the solid phase synthesis. This selectively allows spatial control of the synthesis—the next nucleotide is coupled only to the irradiated areas.

Oligonucleotide synthesis generally involves coupling an activated phosphorous derivative on the 3'-hydroxyl group of a nucleotide with the 5'-hydroxyl group of an oligomer bound to a solid support. Two major chemical methods exist to perform this coupling: the phosphate-triester and phosphoamidite methods (See Gait). Protecting groups of the present invention are suitable for use in either method.

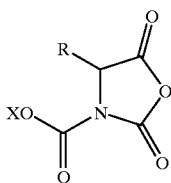
In a preferred embodiment, a photoremovable protecting group is attached to an activated nucleotide on the 5'-hydroxyl group:



where B is the base attached to the sugar ring; R is a hydrogen atom when the sugar is deoxyribose or R is a hydroxyl group when the sugar is ribose; P represents an activated phosphorous group; and X is a photoremovable protecting group. The photoremovable protecting group, X, is preferably NV, NP, PyR, MeNV, MeNP, and the like as described above. The activated phosphorous group, P, is preferably a reactive derivative having a high coupling efficiency, such as a phosphate-triester, phosphoamidite or the like. Other activated phosphorous derivatives, as well as reaction conditions, are well known (See Gait).

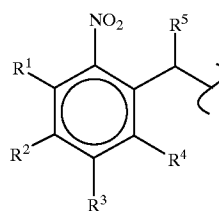
E. Amino Acid N-Carboxy Anhydrides Protected With a Photoremovable Group

During Merrifield peptide synthesis, an activated ester of one amino acid is coupled with the free amino terminus of a substrate-bound oligomer. Activated esters of amino acids suitable for the solid phase synthesis include halo-formate, mixed anhydride, imidazolyl formate, acyl halide, and also includes formation of the activated ester in situ and the use of common reagents such as DCC and the like (See Atherton et al.). A preferred protected and activated amino acid has the general formula:



where R is the side chain of the amino acid and X is a photoremovable protecting group. This compound is a urethane-protected amino acid having a photoremovable protecting group attach to the amine. A more preferred activated amino acid is formed when the photoremovable protecting group has the general formula:

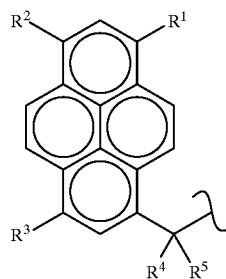
44



where R¹, R², R³, and R⁴ independently are a hydrogen atom, a lower alkyl, aryl, benzyl, halogen, hydroxyl, alkoxy, thiol, thioether, amino, nitro, carboxyl, formate, formamido or phosphido group, or adjacent substituents (i.e., R¹-R², R²-R³, R³-R⁴) are substituted oxygen groups that together form a cyclic acetal or ketal; and R⁵ is a hydrogen atom, an alkoxy, alkyl, hydrogen, halo, aryl, or alkenyl group.

A preferred activated amino acid is formed when the photoremovable protecting group is 6-nitroveratryloxycarbonyl. That is, R¹ and R⁴ are each a hydrogen atom, R² and R³ are each a methoxy group, and R⁵ is a hydrogen atom. Another preferred activated amino acid is formed when the photoremovable group is 6-nitropiperonyl: R¹ and R⁴ are each a hydrogen atom, R² and R³ together form a methylene acetal, and R⁵ is a hydrogen atom. Other protecting groups are possible. Another preferred activated ester is formed when the photoremovable group is methyl-6-nitroveratryl or methyl-6-nitropiperonyl.

Another preferred activated amino acid is formed when the photoremovable protecting group has the general formula:



where R¹, R², and R³ independently are a hydrogen atom, a lower alkyl, aryl, benzyl, halogen, hydroxyl, alkoxy, thiol, thioether, amino, nitro, carboxyl, formate, formamido, sulfanates, sulfido or phosphido group, and R⁴ and R⁵ independently are a hydrogen atom, an alkoxy, alkyl, halo, aryl, hydrogen, or alkenyl group. The resulting compound is a urethane-protected amino acid having a pyrenylmethyloxycarbonyl protecting group attached to the amine. A more preferred embodiment is formed when R¹ through R⁵ are each a hydrogen atom.

The urethane-protected amino acids having a photoremovable protecting group of the present invention are prepared by condensation of an N-protected amino acid with an acylating agent such as an acyl halide, anhydride, chloroformate and the like (See Fuller et al., U.S. Pat. No. 4,946,942 and Fuller et al., *J. Amer. Chem. Soc.* (1990) 112:7414-7416, both herein incorporated by reference for all purposes).

Urethane-protected amino acids having photoremovable protecting groups are generally useful as reagents during solid-phase peptide synthesis, and because of the spatially

US 6,403,320 B1

45

selectivity possible with the photoremovable protecting group, are especially useful for the spatially addressing peptide synthesis. These amino acids are difunctional: the urethane group first serves to activate the carboxy terminus for reaction with the amine bound to the surface and, once the peptide bond is formed, the photoremovable protecting group protects the newly formed amino terminus from further reaction. These amino acids are also highly reactive to nucleophiles, such as deprotected amines on the surface of the solid support, and due to this high reactivity, the solid-phase peptide coupling times are significantly reduced, and yields are typically higher.

1. Example

Light activated formation of a thymidine-cytidine dimer was carried out. A three dimensional representation of a fluorescence scan showing a checkboard pattern generated by the light-directed synthesis of a dinucleotide is shown in FIG. 8. 5'-nitroveratryl thymidine was attached to a synthesis substrate through the 3' hydroxyl group. The nitroveratryl protecting groups were removed by illumination through a 500 nm checkerboard mask. The substrate was then treated with phosphoramidite activated 2'-deoxycytidine. In order to follow the reaction fluorometrically, the deoxycytidine had been modified with an Fmoc protected aminoethyl linker attached to the 5' end. After removal of the Fmoc protecting group with base, the regions which contained the dinucleotide were fluorescently labelled by treatment of the substrate with 1 mM FITC in DMF for one hour.

The three-dimensional representation of the fluorescent intensity data in FIG. 14 clearly reproduces the checkerboard illumination pattern used during photolysis of the substrate. This result demonstrates that oligonucleotides as well as peptides can be synthesized by the light-directed method.

C. Binary Masking

In fact, the means for producing a substrate useful for these techniques are explained in U.S. Ser. No. 07/492,462 (VLSIPS CIP), which is hereby incorporated herein by reference. However, there are various particular ways to optimize the synthetic processes. Many of these methods are described in U.S. Ser. No. 07/624,120 (automated VLSIPS).

Briefly, the binary synthesis strategy refers to an ordered strategy for parallel synthesis of diverse polymer sequences by sequential addition of reagents which may be represented by a reactant matrix, and a switch matrix, the product of which is a product matrix. A reactant matrix is a $1 \times n$ matrix of the building blocks to be added. The switch matrix is all or a subset of the binary numbers from 1 to n arranged in columns. In preferred embodiments, a binary strategy is one in which at least two successive steps illuminate half of a region of interest on the substrate. In most preferred embodiments, binary synthesis refers to a synthesis strategy which also factors a previous addition step. For example, a strategy in which a switch matrix for a masking strategy halves regions that were previously illuminated, illuminating about half of the previously illuminated region and protecting the remaining half (while also protecting about half of previously protected regions and illuminating about half of previously protected regions). It will be recognized that binary rounds may be interspersed with non-binary rounds and that only a portion of a substrate may be subjected to a binary scheme, but will still be considered to be a binary masking scheme within the definition herein. A binary "masking" strategy is a binary synthesis which uses light to remove protective groups from materials for addition of other materials such as nucleotides or amino acids.

In particular, this procedure provides a simplified and highly efficient method for saturating all possible sequences

46

of a defined length polymer. This masking strategy is also particularly useful in producing all possible oligonucleotide sequence probes of a given length.

D. Applications

The technology provided by the present invention has very broad applications. Although described specifically for polynucleotide sequences, similar sequencing, fingerprinting, mapping, and screening procedures can be applied to polypeptide, carbohydrate, or other polymers. In particular, the present invention may be used to completely sequence a given target sequence to subunit resolution. This may be for de novo sequencing, or may be used in conjunction with a second sequencing procedure to provide independent verification. See, e.g., (1988) *Science* 242:1245. For example, a large polynucleotide sequence defined by either the Maxam and Gilbert technique or by the Sanger technique may be verified by using the present invention.

In addition, by selection of appropriate probes, a polynucleotide sequence can be fingerprinted. Fingerprinting is a less detailed sequence analysis which usually involves the characterization of a sequence by a combination of defined features. Sequence fingerprinting is particularly useful because the repertoire of possible features which can be tested is virtually infinite. Moreover, the stringency of matching is also variable depending upon the application. A Southern Blot analysis may be characterized as a means of simple fingerprint analysis.

Fingerprinting analysis may be performed to the resolution of specific nucleotides, or may be used to determine homologies, most commonly for large segments. In particular, an array of oligonucleotide probes of virtually any workable size may be positionally localized on a matrix and used to probe a sequence for either absolute complementary matching, or homology to the desired level of stringency using selected hybridization conditions.

In addition, the present invention provides means for mapping analysis of a target sequence or sequences. Mapping will usually involve the sequential ordering of a plurality of various sequences, or may involve the localization of a particular sequence within a plurality of sequences. This may be achieved by immobilizing particular large segments onto the matrix and probing with a shorter sequence to determine which of the large sequences contain that smaller sequence. Alternatively, relatively shorter probes of known or random sequence may be immobilized to the matrix and a map of various different target sequences may be determined from overlaps. Principles of such an approach are described in some detail by Evans et al. (1989) "Physical Mapping of Complex Genomes by Cosmid Multiplex Analysis," *Proc. Natl. Acad. Sci. USA* 86:5030-5034; Michiels et al. (1987) "Molecular Approaches to Genome Analysis: A Strategy for the Construction of Ordered Overlap Clone Libraries," *CABIOS* 3:203-210; Olsen et al. (1986) "Random-Clone Strategy for Genomic Restriction Mapping in Yeast," *Proc. Natl. Acad. Sci. USA* 83:7826-7830; Craig, et al. (1990) "Ordering of Cosmid Clones Covering the Herpes Simplex Virus Type I (HSV-1) Genome: A Test Case for Fingerprinting by Hybridization," *Nuc. Acids Res.* 18:2653-2660; and Coulson, et al. (1986) "Toward a Physical Map of the Genome of the Nematode *Caenorhabditis elegans*," *Proc. Natl. Acad. Sci. USA* 83:7821-7825; each of which is hereby incorporated herein by reference.

Fingerprinting analysis also provides a means of identification. In addition to its value in apprehension of criminals from whom a biological sample, e.g., blood, has been collected, fingerprinting can ensure personal identification

for other reasons. For example, it may be useful for identification of bodies in tragedies such as fire, flood, and vehicle crashes. In other cases the identification may be useful in identification of persons suffering from amnesia, or of missing persons. Other forensics applications include establishing the identity of a person, e.g., military identification “dog tags”, or may be used in identifying the source of particular biological samples. Fingerprinting technology is described, e.g., in Carrano, et al. (1989) “A High-Resolution, Fluorescence-Based, Semi-automated method for DNA Fingerprinting,” *Genomics* 4: 129–136, which is hereby incorporated herein by reference. See, e.g., table I, for nucleic acid applications, and corresponding applications may be accomplished using polypeptides.

TABLE I

VLSIPS PROJECT IN NUCLEIC ACIDS	
I.	Construction of Chips
II.	Applications
A.	Sequencing
1.	Primary sequencing
2.	Secondary sequencing (sequence checking)
3.	Large scale mapping
4.	Fingerprinting
B.	Duplex/Triplex formation
1.	Antisense
2.	Sequence specific function modulation (e.g. promoter inhibition)
C.	Diagnosis
1.	Genetic markers
2.	Type markers
a.	Blood donors
b.	Tissue transplants
D.	Microbiology
1.	Clinical microbiology
2.	Food microbiology
III.	Instrumentation
A.	Chip machines
B.	Detection
IV.	Software Development
A.	Instrumentation software
B.	Data reduction software
C.	Sequence analysis software

The fingerprinting analysis may be used to perform various types of genetic screening. For example, a single substrate may be generated with a plurality of screening probes, allowing for the simultaneous genetic screening for a large number of genetic markers. Thus, prenatal or diagnostic screening can be simplified, economized, and made more generally accessible.

In addition to the sequencing, fingerprinting, and mapping applications, the present invention also provides means for determining specificity of interaction with particular sequences. Many of these applications were described in U.S. Ser. No. 07/362,901 (VLSIPS parent), U.S. Ser. No. 07/492,462 (VLSIPS CIP), U.S. Ser. No. 07/435,316 (caged biotin parent), and U.S. Ser. No. 07/612,671 (caged biotin CIP).

E. Detection Methods and Apparatus

An appropriate detection method applicable to the selected labeling method can be selected. Suitable labels include radionucleotides, enzymes, substrates, cofactors, inhibitors, magnetic particles, heavy metal atoms, and particularly fluorescers, chemiluminescers, and spectroscopic labels. Patents teaching the use of such labels include U.S. Pat. Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,366,241.

With an appropriate label selected, the detection system best adapted for high resolution and high sensitivity detection may be selected. As indicated above, an optically

detectable system, e.g., fluorescence or chemiluminescence would be preferred. Other detection systems may be adapted to the purpose, e.g., electron microscopy, scanning electron microscopy (SEM), scanning tunneling electron microscopy (STEM), infrared microscopy, atomic force microscopy (AFM), electrical conductance, and image plate transfer.

With a detection method selected, an apparatus for scanning the substrate will be designed. Apparatus, as described in U.S. Ser. No. 07/362,901 (VLSIPS parent); or U.S. Ser. No. 07/492,462 (VLSIPS CIP); or U.S. Ser. No. 07/624,120, (automated VLSIPS), are particularly appropriate. Design modifications may also be incorporated therein.

F. Data Analysis

Data is analyzed by processes similar to those described below in the section describing theoretical analysis. More efficient algorithms will be mathematically devised, and will usually be designed to be performed on a computer. Various computer programs which may more quickly or efficiently make measurement samples and distinguish signal from noise will also be devised. See, particularly, U.S. Ser. No. 07/624,120, (automated VLSIPS).

The initial data resulting from the detection system is an array of data indicative of fluorescent intensity versus location on the substrate. The data are typically taken over regions substantially smaller than the area in which synthesis of a given polymer has taken place. Merely by way of example, if polymers were synthesized in squares on the substrate having dimensions of 500 microns by 500 microns, the data may be taken over regions having dimensions of 5 microns by 5 microns. In most preferred embodiments, the regions over which fluorescence data are taken across the substrate are less than about 1/2 the area of the regions in which individual polymers are synthesized, preferably less than 1/10 the area in which a single polymer is synthesized, and most preferably less than 1/100 the area in which a single polymer is synthesized. Hence, within any area in which a given polymer has been synthesized, a large number of fluorescence data points are collected.

A plot of number of pixels versus intensity for a scan should bear a rough resemblance to a bell curve, but spurious data are observed, particularly at higher intensities. Since it is desirable to use an average of fluorescent intensity over a given synthesis region in determining relative binding affinity, these spurious data will tend to undesirably skew the data.

Accordingly, in one embodiment of the invention the data are corrected for removal of these spurious data points, and an average of the data points is thereafter utilized in determining relative binding efficiency. In general the data are fitted to a base curve and statistically measures are used to remove spurious data.

In an additional analytical tool, various degeneracy reducing analogues may be incorporated in the hybridization probes. Various aspects of this strategy are described, e.g., in Macevicz, S. (1990) PCT publication number WO 90/04652, which is hereby incorporated herein by reference.

II. Theoretical Analysis

The principle of the hybridization sequencing procedure is based, in part, upon the ability to determine overlaps of short segments. The VLSIPS technology provides the ability to generate reagents which will saturate the possible short subsequence recognition possibilities. The principle is most easily illustrated by using a binary sequence, such as a sequence of zeros and ones. Once having illustrated the application to a binary alphabet, the principle may easily be understood to encompass three letter, four letter, five or more. letter, even 20 letter alphabets. A theoretical treatment

of analysis of subsequence information to reconstruction of a target sequence is provided, e.e., in Lysov, Yu., et al. (1988) *Doklady Akadem. Nauk. SSR* 303:1508-1511; Khropko K., et al. (1989) *FEBS Letters* 256:118-122; Pevzner, P. (1989) *J. of Biomolecular Structure and Dynamics* 7:63-69; and Drmanac, R. et al. (1989) *Genomics* 4:114-128; each of which is hereby incorporated herein by reference.

The reagents for recognizing the subsequences will usually be specific for recognizing a particular polymer subsequence anywhere within a target polymer. It is preferable that conditions may be devised which allow absolute discrimination between high fidelity matching and very low levels of mismatching. The reagent interaction will preferably exhibit no sensitivity to flanking sequences, to the subsequence position within the target, or to any other remote structure within the sequence. For polynucleotide sequencing, the specific reagents can be oligonucleotide probes; for polypeptides and carbohydrates, antibodies will be useful reagents. Antibody reagents should also be useful for other types of polymers.

A. Simple n-mer Structure: Theory

1. Simple Two Letter Alphabet: Example

A simple example is presented below of how a sequence of ten digits comprising zeros and ones would be sequenceable using short segments of five digits. For example, consider the sample ten digit sequence:

1010011100.

A VLSIPS substrate could be constructed, as discussed elsewhere, which would have reagents attached in a defined matrix pattern which specifically recognize each of the possible five digit sequences of ones and zeros. The number of possible five digit subsequences is $2^5=32$. The number of possible different sequences 10 digits long is $2^{10}=1,024$. The five contiguous digit subsequences within a ten digit sequence number six, i.e., positioned at digits 1-5, 2-6, 3-7, 4-8, 5-9, and 6-10. It will be noted that the specific order of the digits in the sequence is important and that the order is directional, e.g., running left to right versus right to left. The first five digit sequence contained in the target sequence is 10100. The second is 01001, the third is 10011, the fourth is 00111, the fifth is 01110, and the sixth is 11100.

The VLSIPS substrate would have a matrix pattern of positionally attached reagents which recognize each of the different 5-mer subsequences. Those reagents which recognize each of the 6 contained 5-mers will bind the target, and a label allows the positional determination of where the sequence specific interaction has occurred. By correlation of the position in the matrix pattern, the corresponding bound subsequences can be determined.

In the above-mentioned sequence, six different 5-mer sequences would be determined to be present. They would be:

10100
01001
10011
00111
01110
11100

Any sequence which contains the first five digit sequence, 10100, already narrows the number of possible sequences (e.g., from 1024 possible sequences) which contain it to less than about 192 possible sequences.

This 192 is derived from the observation that with the subsequence 10100 at the far left of the sequence, in positions 1-5, there are only 32 possible sequences.

Likewise, for that particular subsequence in positions 2-6, 3-7, 4-8, 5-9, and 6-10. So, to sum up all of the sequences that could contain 10100, there are 32 for each position and 6 positions for a total of about 192 possible sequences.

However, some of these 10 digit sequences will have been counted twice. Thus, by virtue of containing the 10100 subsequence, the number of possible 10-mer sequences has been decreased from 1024 sequences to less than about 192 sequences.

In this example, not only do we know that sequence contains 10100, but we also know that it contains the second five character sequence, 01001. By virtue of knowing that the sequence contains 10100, we can look specifically to determine whether the sequence contains a subsequence of five characters which contains the four leftmost digits plus a next digit to the left. For example, we would look for a sequence of X1010, but we find that there is none. Thus, we know that the 10100 must be at the left end of the 10-mer. We would also look to see whether the sequence contains the rightmost four digits plus a next digit to the right, e.g., 0100X. We find that the sequence also contains the sequence 01001, and that X is a 1. Thus, we know at least that our target sequence has an overlap of 0100 and has the left terminal sequence 101001.

Applying the same procedure to the second 5-mer, we also know that the sequence must include a sequence of five digits having the sequence 1001Y where Y must be either 0 or 1. We look through the fragments and we see that we have a 10011 sequence within our target, thus Y is also 1. Thus, we would know that our sequence has a sequence of the first seven being 1010011.

Moving to the next 5-mer, we know that there must be sequence of 0011Z, where Z must be either 0 or 1. We look at the fragments produced above and see that the target sequence contains a 00111 subsequence and Z is 1. Thus, we know the sequence must start with 10100111.

The next 5-mer must be of the sequence 0111W where W must be 0 or 1. Again, looking up at the fragments produced, we see that the target sequence contains a 01110 subsequence, and W is a 0. Thus, our sequence to this point is 101001110. We know that the last 5-mer must be either 11100 or 11101. Looking above, we see that it is 11100 and that must be the last of our sequence. Thus, we have determined that our sequence must have been 1010011100.

However, it will be recognized from the example above with the sequences provided therein, that the sequence analysis can start with any known positive probe subsequence. The determination may be performed by moving linearly along the sequence checking the known sequence with a limited number of next positions. Given this possibility, the sequence may be determined, besides by scanning all possible oligonucleotide probe positions, by specifically looking only where the next possible positions would be. This may increase the complexity of the scanning but may provide a longer time span dedicated towards scanning and detecting specific positions of interest relative to other sequence possibilities. Thus, the scanning apparatus could be set up to work its way along a sequence from a given contained oligonucleotide to only look at those positions on the substrate which are expected to have a positive signal.

It is seen that given a sequence, it can be de-constructed into n-mers to produce a set of internal contiguous subsequences. From any given target sequence, we would be able to determine what fragments would result. The hybridization sequence method depends, in part, upon being able to work in the reverse, from a set of fragments of known sequences

US 6,403,320 B1

51

to the full sequence. In simple cases, one is able to start at a single position and work in either or both directions towards the ends of the sequence as illustrated in the example.

The number of possible sequences of a given length increases very quickly with the length of that sequence. Thus, a 10-mer of zeros and ones has 1024 possibilities, a 12-mer has 4096. A 20-mer has over a million possibilities, and a 30-mer has over a billion. However, a given 30-mer has, at most, 26 different internal 5-mer sequences. Thus, a 30 character target sequence having over a million possible sequences can be substantially defined by only 26 different 5-mers. It will be recognized that the probe oligonucleotides will preferably, but need not necessarily, be of identical length, and that the probe sequences need not necessarily be contiguous in that the overlapping subsequences need not differ by only a single subunit. Moreover, each position of the matrix pattern need not be homogeneous, but may actually contain a plurality of probes of known sequence. In addition, although all of the possible subsequence specifications would be preferred, a less than full set of sequences specifications could be used. In particular, although a substantial fraction will preferably be at least about 70%, it may be less than that. About 20% would be preferred, more preferably at least about 30% would be desired. Higher percentages would be especially preferred.

2. Example of Four Letter Alphabet

A four letter alphabet may be conceptualized in at least two different ways from the two letter alphabet. One way, is to consider the four possible values at each position and to analogize in a similar fashion to the binary example each of the overlaps. A second way is to group the binary digits into groups.

Using the first means, the overlap comparisons are performed with a four letter alphabet rather than a two letter alphabet. Then, in contrast to the binary system with 10 positions where $2^{10}=1024$ possible sequences, in a 4-character alphabet with 10 positions, there will actually be $4^{10}=1,048,576$ possible sequences. Thus, the complexity of a four character sequence has a much larger number of possible sequences compared to a two character sequence. Note, however, that there are still only 6 different internal 5-mers. For simplicity, we shall examine a 5 character string with 3 character subsequences. Instead of only 1 and 0, the characters may be designated, e.g., A, C, G, and T. Let us take the sequence GGCTA. The 3-mer subsequences are:

GGC
GCT
CTA

Given these subsequences, there is one sequence, or at most only a few sequences which would produce that combination of subsequences, i.e., GGCTA.

Alternatively, with a four character universe, the binary system can be looked at in pairs of digits. The pairs would be 00, 01, 10, and 11. In this manner, the earlier used sequence 1010011100 is looked at as 10,10,01,11,00. Then the first character of two digits is selected from the possible universe of the four representations 00, 01, 10, and 11. Then a probe would be in an even number of digits, e.g., not five digits, but, three pairs of digits or six digits. A similar comparison is performed and the possible overlaps determined. The 3-pair subsequences are:

52

10, 10, 01
10, 01, 11
01, 11, 00

and the overlap reconstruction produces 10,10,01,11,00.

The latter of the two conceptual views of the 4 letter alphabet provides a representation which is similar to what would be provided in a digital computer. The applicability to a four nucleotide alphabet is easily seen by assigning, e.g., 00 to A, 01 to C, 10 to G, and 11 to T. And, in fact, if such a correspondence is used, both examples for the 4 character sequences can be seen to represent the same target sequence. The applicability of the hybridization method and its analysis for determining the ultimate sequence is easily seen if A is the representation of adenine, C is the representation of cytosine, G is the representation of guanine, and T is the representation of thymine or uracil.

3. Generalization to m-letter Alphabet

This reconstruction process may be applied to polymers of virtually any number of possible characters in the alphabet, and for virtually any length sequence to be sequenced, though limitations, as discussed below, will limit its efficiency at various extremes of length. It will be recognized that the theory can be applied to a large diversity of systems where sequence is important.

For example, the method could be applied to sequencing of a polypeptide. A polypeptide can have any of twenty natural amino acid possibilities at each position. A twenty letter alphabet is amenable to sequencing by this method so long as reagents exist for recognizing shorter subsequences therein. A preferred reagent for achieving that goal would be a set of monoclonal antibodies each of which recognizes a specific three contiguous amino acid subsequence. A complete set of antibodies which recognize all possible subsequences of a given length, e.g., 3 amino acids, and preferably with a uniform affinity, would be $20^3=8000$ reagents.

It will also be recognized that each target sequence which is recognized by the specific reagents need not have homogeneous termini. Thus, fragments of the entire target sequence will also be useful for hybridizing appropriate subsequences. It is, however, preferable that there not be a significant amount of labeled homogeneous contaminating extraneous sequences. This constraint does usually require the purification of the target molecule to be sequenced, but a specific label technique would dispense with a purification requirement if the unlabeled extraneous sequences do not interfere with the labeled sequences.

In addition, conformational effects of target polypeptide folding may, in certain embodiments, be negligible if the polypeptide is fragmented into sufficiently small peptides, or if the interaction is performed under conditions where conformation, but not specific interaction, is disrupted.

B. Complications

Two obvious complications exist with the method of sequence analysis by hybridization. The first results from a probe of inappropriate length while the second relates to internally repeated sequences.

The first obvious complication is a problem which arises from an inappropriate length of recognition sequence, which causes problems with the specificity of recognition. For example, if the recognized sequence is too short, every sequence which is utilized will be recognized by every probe sequence. This occurs, e.g., in a binary system where the probes are each of sequences which occur relatively frequently, e.g., a two character probe for the binary system. Each possible two character probe would be expected to

US 6,403,320 B1

53

appear $\frac{1}{4}$ of the time in every single two character position. Thus, the above sequence example would be recognized by each of the 00, 10, 01, and 11. Thus, the sequence information is virtually lost because the resolution is too low and each recognition reagent specifically binds at multiple sites on the target sequence.

The number of different probes which bind to a target depends on the relationship between the probe length and the target length. At the extreme of short probe length, the just mentioned problem exists of excessive redundancy and lack of resolution. The lack of stability in recognition will also be a problem with extremely short probes. At the extreme of long probe length, each entire probe sequence is on a different position of a substrate. However, a problem arises from the number of possible sequences, which goes up dramatically with the length of the sequence. Also, the specificity of recognition begins to decrease as the contribution to binding by any particular subunit may become sufficiently low that the system fails to distinguish the fidelity of recognition. Mismatched hybridization may be a problem with the polynucleotide sequencing applications, though the fingerprinting and mapping applications may not be so strict in their fidelity requirements. As indicated above, a thirty position binary sequence has over a million possible sequences, a number which starts to become unreasonably large in its required number of different sequences, even though the target length is still very short. Preparing a substrate with all sequence possibilities for a long target may be extremely difficult due to the many different oligomer s which must be synthesized.

The above example illustrates how a long target sequence may be reconstructed with a reasonably small number of shorter subsequences. Since the present day resolution of the regions of the substrate having defined oligomer probes attached to the substrate approaches about 10 microns by 10 microns for resolvable regions, about 10^6 , or 1 million, positions can be placed on a one centimeter square substrate. However, high resolution systems may have particular disadvantages which may be outweighed using the lower density substrate matrix pattern. For this reason, a sufficiently large number of probe sequences can be utilized so that any given target sequence may be determined by hybridization to a relatively small number of probes.

A second complication relates to convergence of sequences to a single subsequence. This will occur when a particular subsequence is repeated in the target sequence. This problem can be addressed in at least two different ways. The first, and simpler way, is to separate the repeat sequences onto two different targets. Thus, each single target will not have the repeated sequence and can be analyzed to its end. This solution, however, complicates the analysis by requiring that some means for cutting at a site between the repeats can be located. Typically a careful sequencer would want to have two intermediate cut points so that the intermediate region can also be sequenced in both directions across each of the cut points. This problem is inherent in the hybridization method for sequencing but can be minimized by using a longer known probe sequence so that the frequency of probe repeats is decreased.

Knowing the sequence of flanking sequences of the repeat will simplify the use of polymerase chain reaction (PCR) or a similar technique to further definitively determine the sequence between sequence repeats. Probes can be made to hybridize to those known sequences adjacent the repeat sequences, thereby producing new target sequences for analysis. See, e.g., Innis et al., (eds.) (1990) *PCR Protocols: A Guide to Methods and Applications*, Academic Press; and

54

methods for synthesis of oligonucleotide probes, see, e.g., Gait (1984) *Oligonucleotide Synthesis: A Practical Approach*, IRL Press, Oxford.

Other means for dealing with convergence problems include using particular longer probes, and using degeneracy reducing analogues, see, e.g., Macevitz, S. (1990) PCT publication number WO 90/04652, which is hereby incorporated herein by reference. By use of stretches of the degeneracy reducing analogues with other probes in particular combinations, the number of probes necessary to fully saturate the possible oligomer probes is decreased. For example, with a stretch of 12-mers having the central 4-mer of degenerate nucleotides, in combination with all of the possible 8-mers, the collection numbers twice the number of possible 8-mers, e.g. $65,536 + 65,536 = 131,072$, but the population provides screening equivalent to all possible 12-mers.

By way of further explanation, all possible oligonucleotide 8-mers may be depicted in the fashion:

N1-N2-N3-N4-N5-N6-N7-N8,

in which there are $4^8 = 65,536$ possible 8-mers. As described in U.S. Ser. No. 07/624,120 (automated VLSIPS), producing all possible 8-mers requires $4 \times 8 = 32$ chemical binary synthesis steps to produce the entire matrix pattern of 65,536 8-mer possibilities. By incorporating degeneracy reducing nucleotides, D's, which hybridize nonselectively to any corresponding complementary nucleotide, new oligonucleotides 12-mers can be made in the fashion:

N1-N2-N3-N4D-D-D-D-N5-N6-N7-N8,

in which there are again, as above, only $4^8 = 65,536$ possible "12-mers", which in reality only have 8 different nucleotides.

However, it can be seen that each possible 12-mer probe could be represented by a group of the two 8-mer types. Moreover, repeats of less than 12 nucleotides would not converge, or cause repeat problems in the analysis. Thus, instead of requiring a collection of probes corresponding to all 12-mers, or $4^{12} = 16,777,216$ different 12-mers, the same information can be derived by making 2 sets of "8-mers" consisting of the typical 8-mer collection of $4^8 = 65,536$ and the "12-mer" set with the degeneracy reducing analogues, also requiring making $4^8 = 65,536$. The combination of the two sets, requires making $65,536 + 65,536 = 131,072$ different molecules, but giving the information of 16,777,216 molecules. Thus, incorporating the degeneracy reducing analogue decreases the number of molecules necessary to get 12-mer resolution by a factor of about 128fold.

C. Non-polynucleotide Embodiments

The above example is directed towards a polynucleotide embodiment. This application is relatively easily achieved because the specific reagents will typically be complementary oligonucleotides, although in certain embodiments other specific reagents may be desired. For example, there may be circumstances where other than complementary base pairing will be utilized. The polynucleotide targets, will usually be single strand, but may be double or triple stranded in various applications. However, a triple stranded specific interaction might be sometimes desired, or a protein or other specific binding molecule may be utilized. For example, various promoter or DNA sequence specific binding proteins might be used, including, e.g., restriction enzyme binding domains, other binding domains, and antibodies. Thus, specific recognition reagents besides oligonucleotides may be utilized.

For other polymer targets, the specific reagents will often be polypeptides. These polypeptides may be protein binding domains from enzymes or other proteins which display specificity for binding. Usually an antibody molecule may

US 6,403,320 B1

55

be used, and monoclonal antibodies may be particularly desired. Classical methods may be applied for preparing antibodies, see, e.g., Harlow and Lane (1988) *Antibodies: A Laboratory Manual* Cold Spring Harbor Press, New York; and Goding (1986) *Monoclonal Antibodies: Principles and Practice* (2d Ed.) Academic Press, San Diego. Other suitable techniques for in vitro exposure of lymphocytes to the antigens or selection of libraries of antibody binding sites are described, e.g., in Huse et al. (1989) *Science* 246:1275-1281; and Ward et al. 91989) *Nature* 341:544-546, each of which is hereby incorporated herein by reference. Unusual antibody production methods are also described, e.g., in Hendricks et al. (1989) *BioTechnology* 7:1271-1274; and Hiatt et al. (1989) *Nature* 342:76-78, each of which is hereby incorporated herein by reference. Other molecules which may exhibit specific binding interaction may be useful for attachment to a VLSIPS substrate by various methods, including the caged biotin methods, see, e.g., U.S. Ser. No. 07/435,316 (caged biotin parent), and U.S. Ser. No. 07/612,671 (caged biotin CIP).

The antibody specific reagents should be particularly useful for the polypeptide, carbohydrate, and synthetic polymer applications. Individual specific reagents might be generated by an automated process to generate the number of reagents necessary to advantageously use the high density positional matrix pattern. In an alternative approach, a plurality of hybridoma cells may be screened for their ability to bind to a VISIPS matrix possessing the desired sequences whose binding specificity is desired. Each cell might be individually grown up and its binding specificity determined by VLSIPS apparatus and technology. An alternative strategy would be to expose the same VLSIPS matrix to a polyclonal serum of high titer. By a successively large volume of serum and different animals, each region of the VLSIPS substrate would have attached to it a substantial number of antibody molecules with specificity of binding. The substrate, with non-covalently bound antibodies could be derivatized and the antibodies transferred to an adjacent second substrate in the matrix pattern in which the antibody molecules had attached to the first matrix. If the sensitivity of detection of binding interaction is sufficiently high, such a low efficiency transfer of antibody molecules may produce a sufficiently high signal to be useful for many purposes, including the sequencing applications.

In another embodiment, capillary forces may be used to transfer the selected reagents to a new matrix, to which the reagents would be positionally attached in the pattern of the recognized sequences. Or, the reagents could be transversely electrophoresed, magnetically transferred, or otherwise transported to a new substrate in their retained positional pattern.

III. Polynucleotide Sequencing

In principle, the making of a substrate having a positionally defined matrix pattern of all possible oligonucleotides of a given length involves a conceptually simple method of synthesizing each and every different possible oligonucleotide, and affixed to a definable position. Oligonucleotide synthesis is presently mechanized and enabled by current technology, see, e.g., U.S. Ser. No. 07/362,901 (VLSIPS parent); U.S. Ser. No. 07/492,462 (VLSIPS CIP); and instruments supplied by Applied Biosystems, Foster City, Calif.

A. Preparation of Substrate Matrix

The production of the collection of specific oligonucleotides used in polynucleotide sequencing may be produced in at least two different ways. Present technology certainly

56

allows production of ten nucleotide oligomers on a solid phase or other synthesizing system. See, e.g., instrumentation provided by Applied Biosystems, Foster City, Calif. Although a single oligonucleotide can be relatively easily made, a large collection of them would typically require a fairly large amount of time and investment. For example, there are $4^{10}=1,048,576$ possible ten nucleotide oligomers. Present technology allows making each and every one of them in a separate purified form though such might be costly and laborious.

Once the desired repertoire of possible oligomer sequences of a given length have been synthesized, this collection of reagents may be individually positionally attached to a substrate, thereby allowing a batchwise hybridization step. Present technology also would allow the possibility of attaching each and every one of these 10-mers to a separate specific position on a solid matrix. This attachment could be automated in any of a number of ways, particularly use of a caged biotin type linking. This would produce a matrix having each of different possible 10-mers.

A batchwise hybridization is much preferred because of its reproducibility and simplicity. An automated process of attaching various reagents to positionally defined sites on a substrate is provided in U.S. Ser. No. 07/492,462 (VLSIPS CIP); U.S. Ser. No. 07/624,120 (automated VLSIPS); and U.S. Ser. No. 07/612,671 (caged biotin CIP), each of which is hereby incorporated herein by reference.

Instead of separate synthesis of each oligonucleotide, these oligonucleotides are conveniently synthesized in parallel by sequential synthetic processes on a defined matrix pattern as provided in U.S. Ser. No. 07/492,462 (VLSIPS CIP); and U.S. Ser. No. 07/624,120, (automated VLSIPS), which are incorporated herein by reference. Here, the oligonucleotides are synthesized stepwise on a substrate at positionally separate and defined positions. Use of photosensitive blocking reagents allows for defined sequences of synthetic steps over the surface of a matrix pattern. By use of the binary masking strategy, the surface of the substrate can be positioned to generate a desired pattern of regions, each having a defined sequence oligonucleotide synthesized and immobilized thereto.

Although the prior art technology can be used to generate the desired repertoire of oligonucleotide probes, an efficient and cost effective means would be to use the VLSIPS technology described in U.S. Ser. No. 07/492,462 (VLSIPS CIP) and U.S. Ser. No. 07/624,120, (automated VLSIPS). In this embodiment, the photosensitive reagents involved in the production of such a matrix are described below.

The regions for synthesis may be very small, usually less than about $100\text{ }\mu\text{m}\times 100\text{ }\mu\text{m}$, more usually less than about $50\text{ }\mu\text{m}\times 50\text{ }\mu\text{m}$. The photolithography technology allows synthetic regions of less than about $10\text{ }\mu\text{m}\times 10\text{ }\mu\text{m}$, about $3\text{ }\mu\text{m}\times 3\text{ }\mu\text{m}$, or less. The detection also may detect such sized regions, though larger areas are more easily and reliably measured.

At a size of about 30 microns by 30 microns, one million regions would take about 11 centimeters square or a single wafer of about 4 centimeters by 4 centimeters. Thus the present technology provides for making a single matrix of that size having all one million plus possible oligonucleotides. Region size are sufficiently small to correspond to densities of at least about 5 regions/cm², 20 regions/cm², 50 regions/cm², 100 regions/cm², and greater, including 300 regions/cm², 1000 regions/cm², 3K regions/cm², 10K regions/cm², 30K regions/cm², 100K regions/cm² 300K regions/cm² or more, even in excess of one million regions/cm².

US 6,403,320 B1

57

Although the pattern of the regions which contain specific sequences is theoretically not important, for practical reasons certain patterns will be preferred in synthesizing the oligonucleotides. The application of binary masking algorithms for generating the pattern of known oligonucleotide probes is described in related U.S. Ser. No. 07/624,120 (automated VLSIPS) which was filed simultaneously with this application. By use of these binary masks, a highly efficient means is provided for producing the substrate with the desired matrix pattern of different sequences. Although the binary masking strategy allows for the synthesis of all lengths of polymers, the strategy may be easily modified to provide only polymers of a given length. This is achieved by omitting steps where a subunit is not attached.

The strategy for generating a specific pattern may take any of a number of different approaches. These approaches are well described in related application U.S. Ser. No. 07/624,120, (automated VLSIPS) and include a number of binary masking approaches which will not be exhaustively discussed herein. However, the binary masking and binary synthesis approaches provide a maximum of diversity with a minimum number of actual synthetic steps.

The length of oligonucleotides used in sequencing applications will be selected on criteria determined to some extent by the practical limits discussed above. For example, if probes are made as oligonucleotides, there will be 65,536 possible eight nucleotide sequences. If a nine subunit oligonucleotide is selected, there are 262,144 possible permutations of sequences. If a ten-mer oligonucleotide is selected, there are 1,048,576 possible permutations of sequences. As the number gets larger, the required number of positionally defined subunits necessary to saturate the possibilities also increases. With respect to hybridization conditions, the length of the matching necessary to converse stability of the conditions selected can be compensated for. See, e.g., Kanehisa, M. (1984) *Nuc. Acids Res.* 12:203-213, which is hereby incorporated herein by reference.

Although not described in detail here, but below for oligonucleotide probes, the VLSIPS technology would typically use a photosensitive protective group on an oligonucleotide. Sample oligonucleotides are shown in FIG. 1. In particular, the photoprotective group on the nucleotide molecules may be selected from a wide variety of positive light reactive groups preferably including nitro aromatic compounds such as o-nitro-benzyl derivatives or benzylsulfonyl. See, e.g., Gait (1984) *Oligonucleotide Synthesis: A Practical Approach*, IRL Press, Oxford, which is hereby incorporated herein by reference. In a preferred embodiment, 6-nitro-veratryl oxycarbonyl (NVOC), 2-nitrobenzyl oxycarbonyl (NBOC), or α,α -dimethyl-dimethoxybenzyl oxycarbonyl (DEZ) is used. Photoremovable protective groups are described in, e.g., Patchornik (1970) *J. Amer. Chem. Soc.* 92:6333-6335; and Amit et al. (1974) *J. Organic Chem.* 39:192-196; each of which is hereby incorporated herein by reference.

A preferred linker for attaching the oligonucleotide to a silicon matrix is illustrated in FIG. 2. A more detailed description is provided below. A photosensitive blocked nucleotide may be attached to specific locations of unblocked prior cycles of attachments on the substrate and can be successively built up to the correct length oligonucleotide probe.

It should be noted that multiple substrates may be simultaneously exposed to a single target sequence where each substrate is a duplicate of one another or where, in combination, multiple substrates together provide the complete or desired subset of possible subsequences. This pro-

58

vides the opportunity to overcome a limitation of the density of positions on a single substrate by using multiple substrates. In the extreme case, each probe might be attached to a single bead or substrate and the beads sorted by whether there is a binding interaction. Those beads which do bind might be encoded to indicate the subsequence specificity of reagents attached thereto.

Then, the target may be bound to the whole collection of beads and those beads that have appropriate specific reagents on them will bind to target. Then a sorting system may be utilized to sort those beads that actually bind the target from those that do not. This may be accomplished by presently available cell sorting devices or a similar apparatus. After the relatively small number of beads which have bound the target have been collected, the encoding scheme may be read off to determine the specificity of the reagent on the bead. An encoding system may include a magnetic system, a shape encoding system, a color encoding system, or a combination of any of these, or any other encoding system. Once again, with the collection of specific interactions that have occurred, the binding may be analyzed for sequence information, fingerprint information, or mapping information.

The parameters of polynucleotide sizes of both the probes and target sequences are determined by the applications and other circumstances. The length of the oligonucleotide probes used will depend in part upon the limitations of the VLSIPS technology to provide the number of desired probes. For example, in an absolute sequencing application, it is often useful to have virtually all of the possible oligonucleotides of a given length. As indicated above, there are 65,536 8-mers, 262,144 9-mers, 1,048,576 10-mers, 4,194,304 11-mers, etc. As the length of the oligomer increases the number of different probes which must be synthesized also increases at a rate of a factor of 4 for every additional nucleotide. Eventually the size of the matrix and the limitations in the resolution of regions in the matrix will reach the point where an increase in number of probes becomes disadvantageous. However, this sequencing procedure requires that the system be able to distinguish, by appropriate selection of hybridization and washing conditions, between binding of absolute fidelity and binding of complementary sequences containing mismatches. On the other hand, if the fidelity is unnecessary, this discrimination is also unnecessary and a significantly longer probe may be used significantly longer probes would typically be useful in fingerprinting or mapping applications.

The length of the probe is selected for a length that it will bind with specificity to possible targets. The hybridization conditions are also very important in that they will determine how close the homology of complementary binding will be detected. In fact, a single target may be evaluated at a number of different conditions to determine its spectrum of specificity for binding particular probes. This may find use in a number of other applications besides the polynucleotide sequencing fingerprinting or mapping. For example, it will be desired to determine the spectrum of binding affinities and specificities of cell surface antigens with binding by particular antibodies immobilized on the substrate surface, particularly under different interaction conditions. In a related fashion, different regions with reagents having differing affinities or levels of specificity may allow such a spectrum to be defined using a single incubation, where various regions, at a given hybridization condition, show the binding affinity. For example, fingerprint probes of various lengths, or with specific defined nonmatches may be used. Unnatural nucleotides or nucleotides exhibiting modified

US 6,403,320 B1

59

specificity of complementary binding are described in greater detail in Macevitz (1990) PCT pub. No. WO 90/04652; and see the section on modified nucleotides in the Sigma chemical company catalogue.

B. Labeling Target Nucleotide

The label used to detect the target sequences will be determined, in part, by the detection methods being applied. Thus, the labeling method and label used are selected in combination with the actual detecting systems being used.

Once a particular label has been selected, appropriate labeling protocols will be applied, as described below for specific embodiments. Standard labeling protocols for nucleic acids are described, e.g., in Sambrook et al.; Kambara, H. et al. (1988) *BioTechnology* 6:816–821; Smith, L. et al. (1985) *Nuc. Acids Res.* 13:2399–2412; for polypeptides, see, e.g., Allen G. (1989) *Sequencing of Proteins and Peptides*, Elsevier, New York, especially chapter 5, and Greenstein and Winitz (1961) *Chemistry of the Amino Acids*, Wiley and Sons, New York. Carbohydrate labeling is described, e.g., in Chaplin and Kennedy (1986) *Carbohydrate Analysis: A Practical Approach*, IRL Press, Oxford. Labeling of other polymers will be performed by methods applicable to them as recognized by a person having ordinary skill in manipulating the corresponding polymer.

In some embodiments, the target need not actually be labeled if a means for detecting where interaction takes place is available. As described below, for a nucleic acid embodiment, such may be provided by an intercalating dye which intercalates only into double stranded segments, e.g., where interaction occurs. See, e.g., Sheldon et al. U.S. Pat. No. 4,582,789.

In many uses, the target sequence will be absolutely homogeneous, both with respect to the total sequence and with respect to the ends of each molecule. Homogeneity with respect to sequence is important to avoid ambiguity. It is preferable that the target sequences of interest not be contaminated with a significant amount of labeled contaminating sequences. The extent of allowable contamination will depend on the sensitivity of the detection system and the inherent signal to noise of the system. Homogeneous contamination sequences will be particularly disruptive of the sequencing procedure.

However, although the target polynucleotide must have a unique sequence, the target molecules need not have identical ends. In fact, the homogeneous target molecule preparation may be randomly sheared to increase the numerical number of molecules. Since the total information content remains the same, the shearing results only in a higher number of distinct sequences which may be labeled and bind to the probe. This fragmentation may give a vastly superior signal relative to a preparation of the target molecules having homogeneous ends. The signal for the hybridization is likely to be dependent on the numerical frequency of the target-probe interactions. If a sequence is individually found on a larger number of separate molecules a better signal will result. In fact, shearing a homogeneous preparation of the target may often be preferred before the labeling procedure is performed, thereby producing a large number of labeling groups associated with each subsequence.

C. Hybridization Conditions

The hybridization conditions between probe and target should be selected such that the specific recognition interaction, i.e., hybridization, of the two molecules is both sufficiently specific and sufficiently stable. See, e.g., Hames and Higgins (1985) *Nucleic Acid Hybridisation: A Practical Approach*, IRL Press, Oxford. These conditions will be dependent both on the specific sequence and often on the

60

guanine and cytosine (GC) content of the complementary hybrid strands. The conditions may often be selected to be universally equally stable independent of the specific sequence involved. This typically will make use of a reagent such as an arylammonium buffer. See, Wood et al. (1985) "Base Composition-independent Hybridization in Tetramethylammonium Chloride: A Method for oligonucleotide screening of Highly Complex Gene Libraries," *Proc. Natl. Acad. Sci. USA*, 82:1585–1588; and Krupov et al. (1989) "An Oligonucleotide Hybridization Approach to DNA Sequencing," *FEBS Letters*, 256:118–122; each of which is hereby incorporated herein by reference. An arylammonium buffer tends to minimize differences in hybridization rate and stability due to GC content. By virtue of the fact that sequences then hybridize with approximately equal affinity and stability, there is relatively little bias in strength or kinetics of binding for particular sequences. Temperature and salt conditions along with other buffer parameters should be selected such that the kinetics of renaturation should be essentially independent of the specific target subsequence or oligonucleotide probe involved. In order to ensure this, the hybridization reactions will usually be performed in a single incubation of all the substrate matrices together exposed to the identical same target probe solution under the same conditions.

Alternatively, various substrates may be individually treated differently. Different substrates may be produced, each having reagents which bind to target subsequences with substantially identical stabilities and kinetics of hybridization. For example, all of the high GC content probes could be synthesized on a single substrate which is treated accordingly. In this embodiment, the arylammonium buffers could be unnecessary. Each substrate is then treated in a manner that the collection of substrates show essentially uniform binding and the hybridization data of target binding to the individual substrate matrix is combined with the data from other substrates to derive the necessary subsequence binding information. The hybridization conditions will usually be selected to be sufficiently specific that the fidelity of base matching will be properly discriminated. Of course, control hybridizations should be included to determine the stringency and kinetics of hybridization.

D. Detection; VLSIPS Scanning

The next step of the sequencing process by hybridization involves labeling of target polynucleotide molecules. A quickly and easily detectable signal is preferred. The VLZ-IPS apparatus is designed to easily detect a fluorescent label, so fluorescent tagging of the target sequence is preferred. Other suitable labels include heavy metal labels, magnetic probes, chromogenic labels (e.g., phosphorescent labels, dyes, and fluorophores) spectroscopic labels, enzyme linked labels, radioactive labels, and labeled binding proteins. Additional labels are described in U.S. Pat. No. 4,366,241, which is incorporated herein by reference.

The detection methods used to determine where hybridization has taken place will typically depend upon the label selected, above. Thus, for a fluorescent label a fluorescent detection step will typically be used. U.S. Ser. No. 07/492,462 (VLSIPS CIP) and U.S. Ser. No. 07/624,120 (automated VLSIPS) describe apparatus and mechanisms for scanning a substrate matrix using fluorescence detection, but a similar apparatus is adaptable for other optically detectable labels.

The detection method provides a positional localization of the region where hybridization has taken place. However, the position is correlated with the specific sequence of the probe since the probe has specifically been attached or synthesized at a defined substrate matrix position. Having

US 6,403,320 B1

61

collected all of the data indicating the subsequences present in the target sequence, this data may be aligned by overlap to reconstruct the entire sequence of the target, as illustrated above.

It is also possible to dispense with actual labeling if some means for detecting the positions of interaction between the sequence specific reagent and the target molecule are available. This may take the form of an additional reagent which can indicate the sites either of interaction, or the sites of lack of interaction, e.g., a negative label. For the nucleic acid embodiments, locations of double strand interaction may be detected by the incorporation of intercalating dyes, or other reagents such as antibody or other reagents that recognize helix formation, see, e.g., Sheldon, et al. (1986) U.S. Pat. No. 4,582,789, which is hereby incorporated herein by reference.

E. Analysis

Although the reconstruction can be performed manually as illustrated above, a computer program will typically be used to perform the overlap analysis. A program may be written and run on any of a large number of different computer hardware systems. The variety of operating systems and languages useable will be recognized by a computer software engineer. Various different languages may be used, e.g., BASIC; C; PASCAL; etc. A simple flow chart of data analysis is illustrated in FIG. 4.

F. Substrate Reuse

Finally, after a particular sequence has been hybridized and the pattern of hybridization analyzed, the matrix substrate should be reusable and readily prepared for exposure to a second or subsequent target polynucleotides. In order to do so, the hybrid duplexes are disrupted and the matrix treated in a way which removes all traces of the original target. The matrix may be treated with various detergents or solvents to which the substrate, the oligonucleotide probes, and the linkages to the substrate are inert. This treatment may include an elevated temperature treatment, treatment with organic or inorganic solvents, modifications in pH, and other means for disrupting specific interaction. Thereafter, a second target may actually be applied to the recycled matrix and analyzed as before.

G. Non-Polynucleotide Aspects

Although the sequencing, fingerprinting, and mapping functions will make use of the natural sequence recognition property of complementary nucleotide sequences, the non-polynucleotide sequences typically require other sequence recognition reagents. These reagents will take the form, typically, of proteins exhibiting binding specificity, e.g., enzyme binding sites or antibody binding sites.

Enzyme binding sites may be derived from promoter proteins, restriction enzymes, and the like. See, e.g., Stryer, L. (1988) *Biochemistry*, W. H. Freeman, Palo Alto. Antibodies will typically be produced using standard procedures, see, e.g., Harlow and Lane (1988) *Antibodies; A Laboratory Manual*, Cold Spring Harbor Press, New York; and Goding (1986) *Monoclonal Antibodies: Principles and Practice*, (2d Ed.) Academic Press, San Diego.

Typically, an antigen, or collection of antigens are presented to an immune system. This may take the form of synthesized short polymers produced by the VLSIPS technology, or by the other synthetic means, or from isolation of natural products. For example, antigen for the polypeptides may be made by the VLSIPS technology, by standard peptide synthesis, by isolation of natural proteins with or without degradation to shorter segments, or by expression of a collection of short nucleic acids of random or defined sequences. See, eg., Tuerk and Gold (1990)

62

Science 249:505-510, for generation of a collection of randomly mutagenized oligonucleotides useful for expression.

The antigen or collection is presented to an appropriate immune system, e.g., to a whole animal as in a standard immunization protocol, or to a collection of immune cells or equivalent. In particular, see Ward et al. (1989) *Nature* 341:544-546; and Huse et al. (1989) *Science* 246:1275-1281, each of which is hereby incorporated herein by reference.

A large diversity of antibodies will be generated, some of which have specificities for the desired sequences. Antibodies may be purified having the desired sequence specificities by isolating the cells producing them. For example, a VLSIPS substrate with the desired antigens synthesized thereon may be used to isolate cells with cell surface reagents which recognize the antigens. The VLSIPS substrate may be used as an affinity reagent to select and recover the appropriate cells. Antibodies from those cells may be attached to a substrate using the caged biotin methodology, or by attaching a targeting molecule, e.g., an oligonucleotide. Alternatively, the supernatants from antibody producing cells can be easily assayed using a VLSIPS substrate to identify the cells producing the appropriate antibodies.

Although cells may be isolated, specific antibody molecules which perform the sequence recognition will also be sufficient. Preferably populations of antibody with a known specificity can be isolated. Supernatants from a large population of producing cells may be passed over a VLSIPS substrate to bind to the desired antigens attached to the substrate. When a sufficient density of antibody molecules are attached, they may be removed by an automated process, preferably as antibody populations exhibiting specificity of binding.

In one particular embodiment, a VLSIPS substrate, e.g., with a large plurality of fingerprint antigens attached thereto, is used to isolate antibodies from a supernatant of a population of cells producing antibodies to the antigens. Using the substrate as an affinity reagent, the antibodies will attach to the appropriate positionally defined antigens. The antibodies may be carefully removed therefrom, preferably by an automated system which retains their homogeneous specificities. The isolated antibodies can be attached to a new substrate in a positionally defined matrix pattern.

In a further embodiment, these spatially separated antibodies may be isolated using a specific targeting method for isolation. In this embodiment, a linker molecule which attaches to a particular portion of the antibody, preferably away from the binding site, can be attached to the antibodies. Various reagents will be used, including staphylococcus protein A or antibodies which bind to domains remote from the binding site. Alternatively, the antibodies in the population, before affinity purification, may be derivatized with an appropriate reagent compatible with new VLSIPS synthesis. A preferred reagent is a nucleotide which can serve as a linker to synthetic VLSIPS steps for synthesizing a specific sequence thereon. Then, by successive VLSIPS cycles, each of the antibodies attached to the defined antigen regions can have a defined oligonucleotide synthesized thereon and corresponding in area to the region of the substrate having each antigen attached. These defined oligonucleotides will be useful as targeting reagents to attach those antibodies possessing the same target sequence specificity at defined positions on a new substrate, by virtue of having bound to the antigen region, to a new VLSIPS substrate having the complementary target oligonucleotides positionally located on it. In this fashion, a VLSIPS substrate

US 6,403,320 B1

63

having the desired antigens attached thereto can be used to generate a second VLSIPS substrate with positionally defined reagents which recognize those antigens.

The selected antigens will typically be selected to be those which define particular functionalities or properties, so as to be useful for fingerprinting and other uses. They will also be useful for mapping and sequencing embodiments.

IV. Fingerprinting

A. General

Many of the procedures and techniques used in the polynucleotide sequencing section are also appropriate for fingerprinting applications. See, e.g., Poustka, et al. (1986) *Cold Spring Harbor Symposia on Quant. Biol.*, vol. LI, 131-139, Cold Spring Harbor Press, New York; which is hereby incorporated herein by reference. The fingerprinting method provided herein is based, in part, upon the ability to positionally localize a large number of different specific probes onto a single substrate. This high density matrix pattern provides the ability to screen for, or detect, a very large number of different sequences simultaneously. In fact, depending upon the hybridization conditions, fingerprinting to the resolution of virtually absolute matching of sequence is possible thereby approaching an absolute sequencing embodiment. And the sequencing embodiment is very useful in identifying the probes useful in further fingerprinting uses. For example, characteristic features of genetic sequences will be identified as being diagnostic of the entire sequence. However, in most embodiments, longer probe and target will be used, and for which light mismatching may not need to be resolved.

B. Preparation of Substrate Matrix

A collection of specific probes may be produced by either of the methods described above in the section on sequencing. Specific oligonucleotide probes of desired lengths may be individually synthesized on a standard oligonucleotide synthesizer. The length of these probes is limited only by the length of the ability of the synthesizer to continue to accurately synthesize a molecule. Oligonucleotides or sequence fragments may also be isolated from natural sources. Biological amplification methods may be coupled with synthetic synthesizing procedures such as, e.g., polymerase chain reaction.

In one embodiment, the individually isolated probes may be attached to the matrix at defined positions. These probe reagents may be attached by an automated process making use of the caged biotin methodology described in U.S. Ser. No. 07/612,671 (caged biotin CIP), or using photochemical reagents, see, e.g., Dattagupta et al. (1985) U.S. Pat. No. 4,542,102 and (1987) U.S. Pat. No. 4,713,326. Each individual purified reagent can be attached individually at specific locations on a substrate.

In another embodiment, the VLSIPS synthesizing technique may be used to synthesize the desired probes at specific positions on a substrate. The probes may be synthesized by successively adding appropriate monomer subunits, e.g., nucleotides, to generate the desired sequences.

In another embodiment, a relatively short specific oligonucleotide is used which serves as a targeting reagent for positionally directing the sequence recognition reagent. For example, the sequence specific reagents having a separate additional sequence recognition segment (usually of a different polymer from the target sequence) can be directed to target oligonucleotides attached to the substrate. By use of non-natural targeting reagents, e.g., unusual nucleotide analogues which pair with other unnatural nucleotide analogues and which do not interfere with natural nucleotide

64

interactions, the natural and non-natural portions can coexist on the same molecule without interfering with their individual functionalities. This can combine both a synthetic and biological production system analogous to the technique for targeting monoclonal antibodies to locations on a VLSIPS substrate at defined positions. Unnatural optical isomers of nucleotides may be useful unnatural reagents subject to similar chemistry, but incapable of interfering with the natural biological polymers. See also, U.S. Ser. No. 07/626,730, (sequencing by synthesis); which is hereby incorporated herein by reference.

After the separate substrate attached reagents are attached to the targeting segment, the two are crosslinked, thereby permanently attaching them to the substrate. Suitable crosslinking reagents are known, see, e.g., Dattagupta et al. (1985) U.S. Pat. No. 4,542,102 and (1987) "Coupling of nucleic acids to solid support by photochemical methods," U.S. Pat. No. 4,713,326, each of which is hereby incorporated herein by reference. Similar linkages for attachment of proteins to a solid substrate are provided, e.g., in Merrifield (1986) *Science* 232:341-347, which is hereby incorporated herein by reference.

C. Labeling Target Nucleotides

The labeling procedures used in the sequencing embodiments will also be applicable in the fingerprinting embodiments. However, since the fingerprinting embodiments often will involve relatively large target molecules and relatively short oligonucleotide probes, the amount of signal necessary to incorporate into the target sequence may be less critical than in the sequencing applications. For example, a relatively long target with a relatively small number of labels per molecule may be easily amplified or detected because of the relatively large target molecule size.

In various embodiments, it may be desired to cleave the target into smaller segments as in the sequencing embodiments. The labeling procedures and cleavage techniques described in the sequencing embodiments would usually also be applicable here.

D. Hybridization Conditions

The hybridization conditions used in fingerprinting embodiments will typically be less critical than for the sequencing embodiments. The reason is that the amount of mismatching which may be useful in providing the fingerprinting information would typically be far greater than that necessary in sequencing uses. For example, Southern hybridizations do not typically distinguish between slightly mismatched sequences. Under these circumstances, important and valuable information may be arrived at with less stringent hybridization conditions while providing valuable fingerprinting information. However, since the entire substrate is typically exposed to the target molecule at one time, the binding affinity of the probes should usually be of approximately comparable levels. For this reason, if oligonucleotide probes are being used, their lengths should be approximately comparable and will be selected to hybridize under conditions which are common for most of the probes on the substrate. Much as in a southern hybridization, the target and oligonucleotide probes are of lengths typically greater than about 25 nucleotides. Under appropriate hybridization conditions, e.g., typically higher salt and lower temperature, the probes will hybridize irrespective of imperfect complementarity. In fact, with probes of greater than, e.g., about fifty nucleotides, the difference in stability of different sized probes will be relatively minor.

Typically the fingerprinting is merely for probing similarity or homology. Thus, the stringency of hybridization can usually be decreased to fairly low levels. See, e.g., Wetmur

US 6,403,320 B1

65

and Davidson (1968) "Kinetics of Renaturation of DNA," *J. Mol. Biol.*, 31:349-370; and Kanehisa, M. (1984) *Nuc. Acids Res.*, 12:203-213.

E. Detection; VLSIPS Scanning

Detection methods will be selected which are appropriate for the selected label. The scanning device need not necessarily be digitized or placed into a specific digital database, though such would most likely be done. For example, the analysis in fingerprinting could be photographic. Where a standardized fingerprint substrate matrix is used, the pattern of hybridizations may be spatially unique and may be compared photographically. In this manner, each sample may have a characteristic pattern of interactions and the likelihood of identical patterns will preferably be such low frequency that the fingerprint pattern indeed becomes a characteristic pattern virtually as unique as an individual's fingertip fingerprint. With a standardized substrate, every individual could be, in theory, uniquely identifiable on the basis of the pattern of hybridizing to the substrate.

Of course, the VLSIPS scanning apparatus may also be useful to generate a digitized version of the fingerprint pattern. In this way, the identification pattern can be provided in a linear string of digits. This sequence could also be used for a standardized identification system providing significant useful medical transferability of specific data. In one embodiment, the probes used are selected to be of sufficiently high resolution to measure the antigens of the major histocompatibility complex, it might even be possible to provide transplantation matching data in a linear stream of data. The fingerprinting data may provide a condensed version, or summary, of the linear genetic data, or any other information data base.

F. Analysis

The analysis of the fingerprint will often be much simpler than a total sequence determination. However, there may be particular types of analysis which will be substantially simplified by a selected group of probes. For example, probes which exhibit particular populational heterogeneity may be selected. In this way, analysis may be simplified and practical utility enhanced merely by careful selection of the specific probes and a careful matrix layout of those probes.

G. Substrate Reuse

As with the sequencing application, the fingerprinting usages may also take advantage of the reusability of the substrate. In this way, the interactions can be disrupted, the substrate treated, and the renewed substrate is equivalent to an unused substrate.

H. Non-polynucleotide Aspects

Besides polynucleotide applications, the fingerprinting analysis may be applied to other polymers, especially polypeptides, carbohydrates, and other polymers, both organic and inorganic. Besides using the fingerprinting method for analyzing a particular polymer, the fingerprinting method may be used to characterize various samples. For example, a cell or population of cells may be tested for their expression of specific antigens or their mRNA sequence intent. For example, a T-cell may be classified by virtue of its combination of expressed surface antigens. With specific reagents which interact with these antigens, a cell or a population of cells or a lysed cell may be exposed to a VLSIPS substrate. The biological sample may be classified or characterized by analyzing the pattern of specific interaction. This may be applicable to a cell or tissue type, to the expressed messenger RNA population expressed by a cell to the genetic content of a cell, or to virtually any sample which can be classified and/or identified by its combination of specific molecular properties.

66

The ability to generate a high density means for screening the presence or absence of specific interactions allows for the possibility of screening for, if not saturating, all of a very large number of possible interactions. This is very powerful in providing the means for testing the combinations of molecular properties which can define a class of samples. For example, a species of organism may be characterized by its DNA sequences, e.g., a genetic fingerprint. By using a fingerprinting method, it may be determined that all members of that species are sufficiently similar in specific sequences that they can be easily identified as being within a particular group. Thus, newly defined classes may be resolved by their similarity in fingerprint patterns. Alternatively, a non-member of that group will fail to share those many identifying characteristics. However, since the technology allows testing of a very large number of specific interactions, it also provides the ability to more finely distinguish between closely related different cells or samples. This will have important applications in diagnosing viral, bacterial, and other pathological or nonpathological infections.

In particular, cell classification may be defined by any of a number of different properties. For example, a cell class may be defined by its DNA sequences contained therein. This allows species identification for parasitic or other infections. For example, the human cell is presumably genetically distinguishable from a monkey cell, but different human cells will share many genetic markers. At higher resolution, each individual human genome will exhibit unique sequences that can define it as a single individual.

Likewise, a developmental stage of a cell type may be definable by its pattern of expression of messenger RNA. For example, in particular stages of cells, high levels of ribosomal RNA are found whereas relatively low levels of other types of messenger RNAs may be found. The high resolution distinguishability provided by this fingerprinting method allows the distinction between cells which have relatively minor differences in its expressed mRNA population. Where a pattern is shown to be characteristic of a stage, a stage may be defined by that particular pattern of messenger RNA expression.

In a similar manner, the antigenic determinants found on a protein may very well define the cell class. For example, immunological T-cells are distinguishable from B-cells because, in part, the cell surface antigens on the cell types are distinguishable. Different T-cell subclasses can be also distinguished from one another by whether they contain particular T-cell antigens. The present invention provides the possibility for high resolution testing of many different interactions simultaneously, and the definition of new cell types will be possible.

The high resolution VLSIPS substrate may also be used as a very powerful diagnostic tool to test the combination of presence, of a plurality of different assays from a biological sample. For example, a cancerous condition may be indicated by a combination of various different properties found in the blood. For example, a cancerous condition may be indicated by a combination of expression of various soluble antigens found in the blood along with a high number of various cellular antigens found on lymphocytes and/or particular cell degradation products. With a substrate as provided herein, a large number of different features can be simultaneously performed on a biological sample. In fact, the high resolution of the test will allow more complete characterization of parameters which define particular diseases. Thus, the power of diagnostic tests may be limited by the extent of statistical correlation with a particular condi-

US 6,403,320 B1

67

tion rather than with the number of antigens or interactions which are tested. The present invention provides the means to generate this large universe of possible reagents and the ability to actually accumulate that correlative data.

In another embodiment, a substrate as provided herein may be used for genetic screening. This would allow for simultaneous screening of thousands of genetic markers. As the density of the matrix is increased, many more molecules can be simultaneously tested. Genetic screening then becomes a simpler method as the present invention provides the ability to screen for thousands, tens of thousands, and hundreds of thousands, even millions of different possible genetic features. However, the number of high correlation genetic markers for conditions numbers only in the hundreds. Again, the possibility for screening a large number of sequences provides the opportunity for generating the data which can provide correlation between sequences and specific conditions or susceptibility. The present invention provides the means to generate extremely valuable correlations useful for the genetic detection of the causative mutation leading to medical conditions. In still another embodiment, the present invention would be applicable to distinguishing two individuals having identical genetic compositions. The antibody population within an individual is dependent both on genetic and historical factors. Each individual experiences a unique exposure to various infectious agents, and the combined antibody expression is partly determined thereby. Thus, individuals may also be fingerprinted by their immunological content, either of actively expressed antibodies, or their immunological memory. Similar sorts of immunological and environmental histories may be useful for fingerprinting, perhaps in combination with other screening properties. In particular, the present invention may be useful for screening allergic reactions or susceptibilities, a simple IgE specificity test may be useful in determining a spectrum of allergies.

With the definition of new classes of cells, a cell sorter will be used to purify them. Moreover, new markers for defining that class of cells will be identified. For example, where the class is defined by its RNA content, cells may be screened by antisense probes which detect the presence or absence of specific sequences therein. Alternatively, cell lysates may provide information useful in correlating intracellular properties with extracellular markers which indicate functional differences. Using standard cell sorter technology with a fluorescence or labeled antisense probe which recognizes the internal presence of the specific sequences of interest, the cell sorter will be able to isolate a relatively homogeneous population of cells possessing the particular marker. Using successive probes the sorting process should be able to select for cells having a combination of a large number of different markers.

In a nonpolynucleotide embodiment, cells may be defined by the presence of other markers. The markers may be carbohydrates, proteins, or other molecules. Thus, a substrate having particular specific reagents, e.g., antibodies, attached to it should be able to identify cells having particular patterns of marker expression. Of course, combinations of these made be utilized and a cell class may be defined by a combination of its expressed mRNA, its carbohydrate expression, its antigens, and other properties. This fingerprinting should be useful in determining the physiological state of a cell or population of cells.

Having defined a cell type whose function or properties are defined by the reagents attachable to a VLSIPS substrate, such as cellular antigens, these structural manifestations of function may be used to sort cells to generate a relatively

68

homogeneous population of that class of cells. Standard cell sorter technology may be applied to purify such a population, see, e.g.; Dangl, J. and Herzenberg (1982) "Selection of hybridomas and hybridoma variants using the fluorescence activated cell sorter," *J. Immunological Methods* 52:1-14; and Becton Dickinson, Fluorescence Activated Cell Sorter Division, San Jose, Calif., and Coulter Diagnostics, Hialeah, Florida.

With the fingerprinted method as in identification means arises from mosaism problems in an organism. A mosaic organism is one whose genetic content in different cells is significantly different. Various clonal populations should have similar genetic fingerprints, though different clonal populations may have different genetic contents. See, for example, Suzuki et al. *An Introduction to Genetic Analysis* (4th Ed.), Freeman and Co., New York, which is hereby incorporated herein by reference. However, this problem should be a relatively rare problem and could be more carefully evaluated with greater experience using the fingerprinting methods.

The invention will also find use in detecting changes, both genetic and antigenic, e.g., in a rapidly "evolving" protozoa infection, or similarly changing organism.

V. Mapping

A. General

The use of the present invention for mapping parallels its use for fingerprinting and sequencing. Where a polymer is a linear molecule, the mapping provides the ability to locate particular segments along the length of the polymer. Branched polymers can be treated as a series of individual linear polymers. The mapping provides the ability to locate, in a relative sense, the order of various subsequences. This may be achieved using at least two different approaches.

The first approach is to take the large sequence and fragment it at specific points. The fragments are then ordered and attached to a solid substrate. For example, the clones resulting from a chromosome walking process may be individually attached to the substrate by methods, e.g., caged biotin techniques, indicated earlier. Segments of unknown map position will be exposed to the substrate and will hybridize to the segment which contains that particular sequence. This procedure allows the rapid determination of a number of different labeled segments, each mapping requiring only a single hybridization step once the substrate is generated. The substrate may be regenerated by removal of the interaction, and the next mapping segment applied.

In an alternative method, a plurality of subsequences can be attached to a substrate. Various short probes may be applied to determine which segments may contain particular overlaps. The theoretical basis and a description of this mapping procedure is contained in, e.g., Evans et al. 1989 "Physical Mapping of Complex Genomes by Cosmid Multiplex Analysis," *Proc. Natl. Acad. Sci. USA* 86:5030-5034, and other references cited above in the Section labeled "Overall Description." Using this approach, the details of the mapping embodiment are very similar to those used in the fingerprinting embodiment.

B. Preparation of Substrate Matrix

The substrate may be generated in either of the methods generally applicable in the sequencing and fingerprinting embodiments. The substrate may be made either synthetically, or by attaching otherwise purified probes or sequences to the matrix. The probes or sequences may be derived either from synthetic or biological means. As indicated above, the solid phase substrate synthetic methods may be utilized to generate a matrix with positionally defined sequences. In the mapping embodiment, the impor-

US 6,403,320 B1

69

tance of saturation of all possible subsequences of a preselected length is far less important than in the sequencing embodiment, but the length of the probes used may be desired to be much longer. The processes for making a substrate which has longer oligonucleotide probes should not be significantly different from those described for the sequencing embodiments, but the optimization parameters may be modified to comply with the mapping needs.

C. Labeling

The labeling methods will be similar to those applicable in sequencing and fingerprinting embodiments. Again, the target sequences may be desired to be fragmented.

D. Hybridization/Specific Interaction

The specificity of interaction between the targets and probe would typically be closer to those used for fingerprinting embodiments, where homology is more important than absolute distinguishability of high fidelity complementary hybridization. Usually, the hybridization conditions will be such that merely homologous segments will interact and provide a positive signal. Much like the fingerprinting embodiment, it may be useful to measure the extent of homology by successive incubations at higher stringency conditions. Or, a plurality of different probes, each having various levels of homology may be used. In either way, the spectrum of homologies can be measured.

Where non-nucleic acid hybridization is involved, the specific interactions may also be compared in a fingerprint-like manner. The specific reagents may have less specificity, e.g., monoclonal antibodies which recognize a broader spectrum of sequences may be utilized relative to a sequencing embodiment. Again, the specificity of interaction may be measured under various conditions of increasing stringency to determine the spectrum of matching across the specific probes selected, or a number of different stringency reagents may be included to indicate the binding affinity.

E. Detection

The detection methods used in the mapping procedure will be virtually identical to those used in the fingerprinting embodiment. The detection methods will be selected in combination with the labeling methods.

F. Analysis

The analysis of the data in a mapping embodiment will typically be somewhat different from that in fingerprinting. The fingerprinting embodiment will test for the presence or absence of specific or homologous segments. However, in the mapping embodiment, the existence of an interaction is coupled with some indication of the location of the interaction. The interaction is mapped in some manner to the physical polymer sequence. Some means for determining the relative positions of different probes is performed. This may be achieved by synthesis of the substrate in pattern, or may result from analysis of sequences after they have been attached to the substrate.

For example, the probes may be randomly positioned at various locations on the substrate. However, the relative positions of the various reagents in the original polymer may be determined by using short fragments, e.g., individually, as target molecules which determine the proximity of different probes. By an automated system of testing each different short fragment of the original polymer, coupled with proper analysis, it will be possible to determine which probes are adjacent one another on the original target sequence and correlate that with positions on the matrix. In this way, the matrix is useful for determining the relative locations of various new segments in the original target molecule. This sort of analysis is described in Evans, and the related references described above.

70

G. Substrate Reuse

The substrate should be reusable in the manner described in the fingerprinting section. The substrate is renewed by removal of the specific interactions and is washed and prepared for successive cycles of exposure to new target sequences.

H. Non-polynucleotide Aspects

The mapping procedure may be used on other molecules than polynucleotides. Although hybridization is one type of specific interaction which is clearly useful for use in this mapping embodiment, antibody reagents may also be very useful. In the same way that polypeptide sequencing or other polymers may be sequenced by the reagents and techniques described in the sequencing section and fingerprinting section, the mapping embodiment may also be used similarly.

In another form of mapping, as described above in the fingerprinting section, the developmental map of a cell or biological system may be measured using fingerprinting type technology. Thus, the mapping may be along a temporal dimension rather than along a polymer dimension. The mapping or fingerprinting embodiments may also be used in determining the genetic rearrangements which may be genetically important, as in lymphocyte and B-cell development. In another example, various rearrangements or chromosomal dislocations may be tested by either the fingerprinting or mapping methods. These techniques are similar in many respects and the fingerprinting and mapping embodiments may overlap in many respects.

VI. Additional Screening and Applications

A. Specific Interactions

As originally indicated in the parent filing of VISIPS, the production of a high density plurality of spatially segregated polymers provides the ability to generate a very large universe or repertoire of individually and distinct sequence possibilities. As indicated above, particular oligonucleotides may be synthesized in automated fashion at specific locations on a matrix. In fact, these oligonucleotides may be used to direct other molecules to specific locations by linking specific oligonucleotides to other reagents which are in batch exposed to the matrix and hybridized in a complementary fashion to only those locations where the complementary oligonucleotide has been synthesized on the matrix. This allows for spatially attaching a plurality of different reagents onto the matrix instead of individually attaching each separate reagent at each specific location. Although the caged biotin method allows the automated attachment, the speed of the caged biotin attachment process is relatively slow and requires a separate reaction for each reagent being attached. By use of the oligonucleotide method, the specificity of position can be done in an automated and parallel fashion. As each reagent is produced, instead of directly attaching each reagent at each desired position, the reagent may be attached to a specific desired complementary oligonucleotide which will ultimately be specifically directed toward locations on the matrix having a complementary oligonucleotide attached thereto.

In addition, the technology allows screening for specificity of interaction with particular reagents. For example, the oligonucleotide sequence specificity of binding of a potential reagent may be tested by presenting to the reagent all of the possible subsequences available for binding. Although secondary or higher order sequence specific features might not be easily screenable using this technology, it does provide a convenient, simple, quick, and thorough screen of interactions between a reagent and its target recognition sequences. See, e.g., Pfeifer et al. (1989) *Science* 246:810-812.

US 6,403,320 B1

71

For example, the interaction of a promoter protein with its target binding sequence may be tested for many different, or all, possible binding sequences. By testing the strength of interactions under various different conditions, the interaction of the promoter protein with each of the different potential binding sites may be analyzed. The spectrum of strength of interactions with each different potential binding site may provide significant insight into the types of features which are important in determining specificity.

An additional example of a sequence specific interaction between reagents is the testing of binding of a double stranded nucleic acid structure with a single stranded oligonucleotide. Often, a triple stranded structure is produced which has significant aspects of sequence specificity. Testing of such interactions with either sequences comprising only natural nucleotides, or perhaps the testing of nucleotide analogs may be very important in screening for particularly useful diagnostic or therapeutic reagents. See, e.g., Häner and Dervan (1990) *Biochemistry* 29:9761-6765, and references therein.

B. Sequence Comparisons

Once a gene is sequenced, the present invention provides means to compare alleles or related sequences to locate and identify differences from the control sequence. This would be extremely useful in further analysis of genetic variability at a specific gene locus.

C. Categorizations

As indicated above in the fingerprinting and mapping embodiments, the present invention is also useful to define specific stages in the temporal sequence of cells, e.g., development, and the resulting tissues within an organism. For example, the developmental stage of a cell, or population of cells, can be dependent upon the expression of particular messenger RNAs or cellular antigens. The screening procedures provided allow for high resolution definition of new classes of cells. In addition, the temporal development of particular cells will be characterized by the presence or expression of various mRNAs. Means to simultaneously screen a plurality or very large number of different sequences as provided. The combination of different markers made available dramatically increases the ability to distinguish fairly closely related cell types. Other markers may be combined with markers and methods made available herein to define new classifications of biological samples, e.g., based upon new combinations of markers.

The presence or absence of particular marker sequences will be used to define temporal developmental stages. Once the stages are defined, fairly simple methods can be applied to actually purify those particular cells. For example, antisense probes or recognition reagents may be used with a cell sorter to select those cells containing or expressing the critical markers. Alternatively, the expression of those sequences may result in specific antigens which may also be used in defining cell classes and sorting those cells away from others. In this way, for example, it should be possible to select a class of omnipotent immune system cells which are able to completely regenerate a human immune system. Based upon the cellular classes defined by the parameters made available by this technology, purified classes of cells having identifiable differences, structural or functional, are made available.

In an alternative embodiment, a plurality of antigens or specific binding proteins attached to the substrate may be used to define particular cell types. For example, subclasses of T-cells are defined, in part, upon the combination of expressed cell surface antigens. The present invention allows for the simultaneous screening of a large plurality of

72

different antigens together. Thus, higher resolution classification of different T-cell subclasses becomes possible and, with the definitions and functional differences which correlate with those antigenic or other parameters, the ability to purify those cell types becomes available. This is applicable not only to T-cells, lymphocyte cells, or even to freely circulating cells. Many of the cells for which this would be most useful will be immobile cells found in particular tissues or organs. Tumor cells will be diagnosed or detected using these fingerprinting techniques. Coupled with a temporal change in structure, developmental classes may also be selected and defined using these technologies. The present invention also provides the ability not only to define new classes of cells based upon functional or structural differences, but it also provides the ability to select or purify populations of cells which share these particular properties. Standard cell sorting procedures using antibody markers may be used to detect extracellular features. Intracellular features would also be amendable by introducing the label reagents into the cell. In particular, antisense DNA or RNA molecules may be introduced into a cell to detect RNA sequences therein. See, e.g., Weintraub (1990) *Scientific American* 262:40-46.

D. Statistical Correlations

In an additional embodiment, the present invention also allows for the high resolution correlation of medical conditions with various different markers. For example, the present technology, when applied to amniocentesis or other genetic screening methods, typically screen for tens of different markers at most. The present invention allows simultaneous screening for tens, hundreds, thousands, tens of thousands, hundreds of thousands, and even millions of different genetic sequences. Thus, applying the fingerprinting methods of the present invention to a sufficiently large population allows detailed statistical analysis to be made, thereby correlating particular medical conditions with particular markers, typically antigenic or genetic. Tumor specific antigens will be identified using the present invention.

Various medical conditions may be correlated against an enormous database of the sequences within an individual. Genetic propensities and correlations then become available and high resolution genetic predictability and correlation become much more easily performed. With the enormous data base, the reliability of the predictions also is better tested. Particular markers which are partially diagnostic of particular medical conditions or medical susceptibilities will be identified and provide direction in further studies and more careful analysis of the markers involved. Of course, as indicated above in the sequencing embodiment, the present invention will find much use in intense sequencing projects. For example, sequencing of the entire human genome in the human genome project will be greatly simplified and enabled by the present invention.

VI. Formation of Substrate

The substrate is provided with a pattern of specific reagents which are positionally localized on the surface of the substrate. This matrix of positions is defined by the automated system which produces the substrate. The instrument will typically be one similar to that described in U.S. Ser. No. 07/492,462 (VLSIPS CIP), and U.S. Ser. No. 07/624,120, (automated VLSIPS). The instrumentation described therein is directly applicable to the applications used here. In particular, the apparatus comprises a substrate, typically a silicon containing substrate, on which positions on the surface may be defined by a coordinate system of positions. These positions can be individually addressed or detected by the VLSIPS apparatus.

US 6,403,320 B1

73

Typically, the VLSIPS apparatus uses optical methods used in semiconductor fabrication applications. In this way, masks may be used to photo-activate positions for attachment or synthesis of specific sequences on the substrate. These manipulations may be automated by the types of apparatus described in U.S. Ser. No. 07/462,492 (VLSIPS CIP) and U.S. Ser. No. 07/624,120, (automated VLSIPS).

Selectively removable protecting groups allow creation of well defined areas of substrate surface having differing reactivities. Preferably, the protecting groups are selectively removed from the surface by applying a specific activator, such as electromagnetic radiation of a specific wavelength and intensity. More preferably, the specific activator exposes selected areas of surface to remove the protecting groups in the exposed areas.

Protecting groups of the present invention are used in conjunction with solid phase oligomer syntheses, such as peptide syntheses using natural or unnatural amino acids, nucleotide syntheses using deoxyribonucleic and ribonucleic acids, oligosaccharide syntheses, and the like. In addition to protecting the substrate surface from unwanted reaction, the protecting groups block a reactive end of the monomer to prevent self-polymerization. For instance, attachment of a protecting group to the amino terminus of an activated amino acid, such as the N-hydroxysuccinimide-activated ester of the amino acid prevents the amino terminus of one monomer from reacting with the activated ester portion of another during peptide synthesis.

Alternatively, the protecting group may be attached to the carboxyl group of an amino acid to prevent reaction at this site. Most protecting groups can be attached to either the amino or the carboxyl group of an amino acid, and the nature of the chemical synthesis will dictate which reactive group will require a protecting group. Analogously, attachment of a protecting group to the 5'-hydroxyl group of a nucleoside during synthesis using for example, phosphate-triester coupling chemistry, prevents the 5'-hydroxyl of one nucleoside from reacting with the 3'-activated phosphate-triester of another.

Regardless of the specific use, protecting groups are employed to protect a moiety on a molecule from reacting with another reagent. Protecting groups of the present invention have the following characteristics: they prevent selected reagents from modifying the group to which they are attached; they are stable (that is, they remain attached) to the synthesis reaction conditions; they are removable under conditions that do not adversely affect the remaining structure; and once removed, do not react appreciably with the surface or surface-bound oligomer. The selection of a suitable protecting group will depend, of course, on the chemical nature of the monomer unit and oligomer, as well as the specific reagents they are to protect against.

In a preferred embodiment, the protecting groups will be photoactivatable. The properties and uses of photoreactive protecting compounds have been reviewed. See, McCray et al., *Ann. Rev. of Biophys. and Biophys. Chem.* (1989) 18:239-270, which is incorporated herein by reference. Preferably, the photosensitive protecting groups will be removable by radiation in the ultraviolet (UV) or visible portion of the electromagnetic spectrum. More preferably, the protecting groups will be removable by radiation in the near UV or visible portion of the spectrum. In some embodiments, however, activation may be performed by other methods such as localized heating, electron beam lithography, laser pumping, oxidation or reduction with micro electrodes, and the like. Sulfonyl compounds are

74

suitable reactive groups for electron beam lithography. Oxidative or reductive removal is accomplished by exposure of the protecting group to an electric current source, preferably using microelectrodes directed to the predefined regions of the surface which are desired for activation. A more detailed description of these protective groups is provided in U.S. Ser. No. 07/624,120, (automated VLSIPS), which is hereby incorporated herein by reference.

The density of reagents attached to a silicon substrate may be varied by standard procedures. The surface area for attachment of reagents may be increased by modifying the silicon surface. For example, a matte surface may be machined or etched on the substrate to provide more sites for attachment of the particular reagents. Another way to increase the density of reagent binding sites is to increase the derivitization density of the silicon. Standard procedures for achieving this are described, below.

One method to control the derivatization density is to highly derivatize the substrate with photochemical groups at high density. The substrate is then photolyzed for various predetermined times, which photoactivate the groups at a measurable rate, and react then with a capping reagent. By this method, the density of linker groups may be modulated by using a desired time and intensity of photoactivation.

In many applications, the number of different sequences which may be provided may be limited by the density and the size of the substrate on which the matrix pattern is generated. In situations where the density is insufficiently high to allow the screening of the desired number of sequences, multiple substrates may be used to increase the number of sequences tested. Thus, the number of sequences tested may be increased by using a plurality of different substrates. Because the VLSIPS apparatus is almost fully automated, increasing the number of substrates does not lead to a significant increase in the number of manipulations which must be performed by humans. This again leads to greater reproducibility and speed in the handling of these multiple substrates.

A. Instrumentation

The concept of using VLSIPS generally allows a pattern or a matrix of reagents to be generated. The procedure for making the pattern is performed by any of a number of different methods. An apparatus and instrumentation useful for generating a high density VLSIPS substrate is described in detail in U.S. Ser. No. 07/492,462 (VLSIPS CIP) and U.S. Ser. No. 07/624,120, (automated VLSIPS).

B. Binary Masking

The details of the binary masking are described in an accompanying application filed simultaneously with this, U.S. Ser. No. 07/624,120 (automated VLSIPS) whose specification is incorporated herein by reference.

For example, the binary masking technique allows for producing a plurality of sequences based on the selection of either of two possibilities at any particular location. By a series of binary masking steps, the binary decision may be the determination, on a particular synthetic cycle, whether or not to add any particular one of the possible subunits. By treating various regions of the matrix pattern in parallel, the binary masking strategy provides the ability to carry out spatially addressable parallel synthesis.

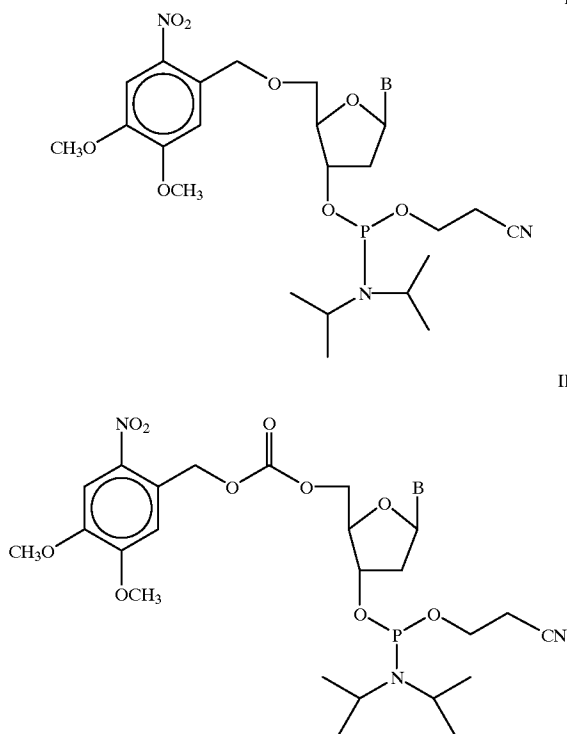
C. Synthetic Methods

The synthetic methods in making a substrate are described in the parent application, U.S. Ser. No. 07/492,462. The

US 6,403,320 B1

75

construction of the matrix pattern on the substrate will typically be generated by the use of photosensitive reagents. By use of photo-lithographic optical methods, particular segments of the substrate can be irradiated with light to activate or deactivate blocking agents, e.g., to protect or deprotect particular chemical groups. By an appropriate sequence of photo-exposure steps at appropriate times with appropriate masks and with appropriate reagents, the substrates can have known polymers synthesized at positionally defined regions on the substrate. Methods for synthesizing various substrates are described in U.S. Ser. No. 07/492,462 (VLSIPS CIP) and U.S. Ser. No. 07/624,120, (automated VLSIPS). By a sequential series of these photo-exposure and reaction manipulations, a defined matrix pattern of known sequences may be generated, and is typically referred to as a VLSIPS substrate. In the nucleic acid synthesis embodiment, nucleosides used in the synthesis of DNA by photolytic methods will typically be one of the two forms shown below:

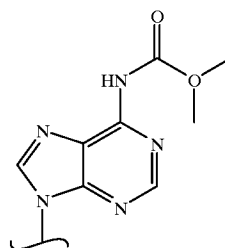


B=Adenine, Cytosine, Guanine, or Thymine

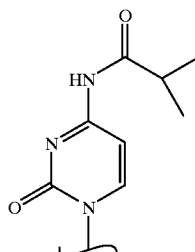
In I, the photolabile group at the 5' position is abbreviated NV (nitroveratryl) and in II, the group is abbreviated NVOC (nitroveratryl oxycarbonyl). Although not shown in FIG. C the bases (adenine, cytosine, and guanine) contain exocyclic NH₂ groups which must be protected during DNA synthesis. Thymine contains no exocyclic NH₂ and therefore requires no protection. The standard protecting groups for these anines are shown below:

76

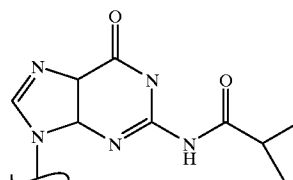
Adenine (A)



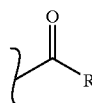
Cytosine (C)



Guanine (G)



Other amides of the general formula

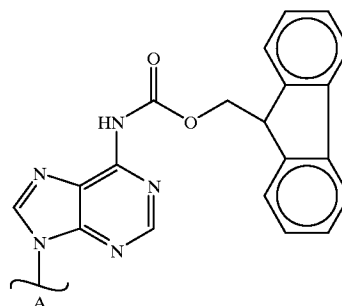


R = alkyl
aryl

where R may be alkyl or aryl have been used.

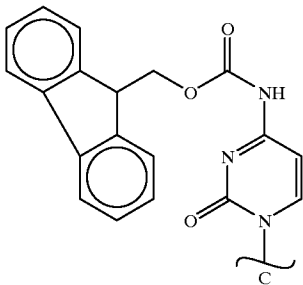
Another type of protecting group Fmoc (9-fluorenyl methoxycarbonyl) is currently being used to protect the exocyclic amines of the three bases:

Adenine (A)



US 6,403,320 B1

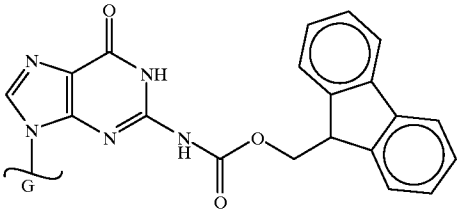
77
-continued



Cytosine (C) 5

10

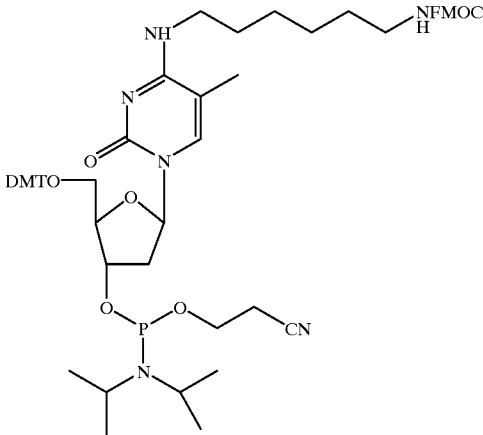
78
-continued Guanine (G)



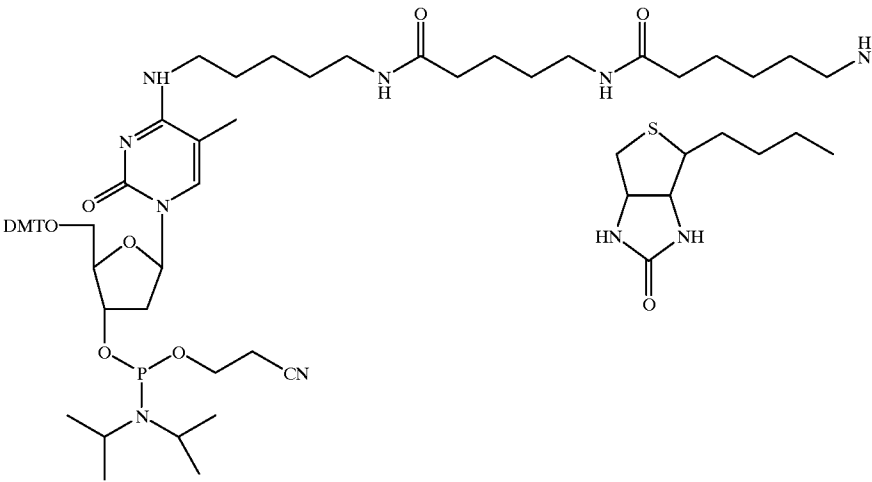
15

The advantage of the Fmoc group is that it is removed under mild conditions (dilute organic bases) and can be used for all three bases. The amide protecting groups require more harsh conditions to be removed (NH₃/MeOH with heat).
Nucleosides used as 5'-OH probes, useful in verifying correct VLSIPS synthetic function, have been the following:

III



IV



US 6,403,320 B1

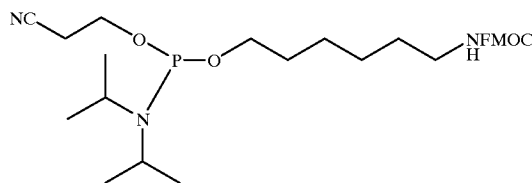
79

These compounds are used to detect where on a substrate photolysis has occurred by the attachment of either III or V to the newly generated 5'-OH. In the case of III, after the phosphate attachment is made, the substrate is treated with a dilute base to remove the Fmoc group. The resulting amine can be reacted with FITC and the substrate examined

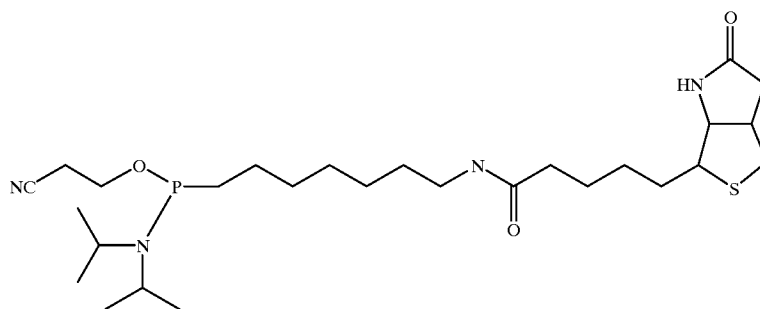
80

by fluorescence microscopy. This indicates the proper generation of a 5'-OH. In the case of compound IV, after the phosphate attachment is made, the substrate is treated with FITC labeled streptavidin and the substrate again may be examined by fluorescence microscopy. Other probes, although not nucleoside based, have included the following:

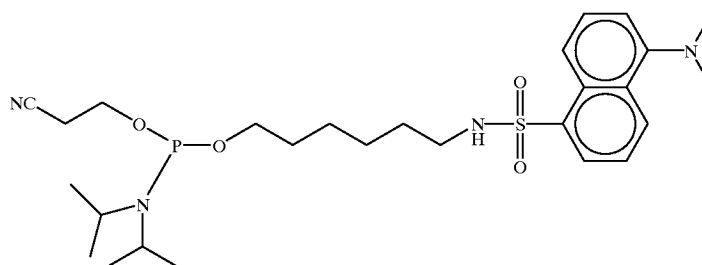
V



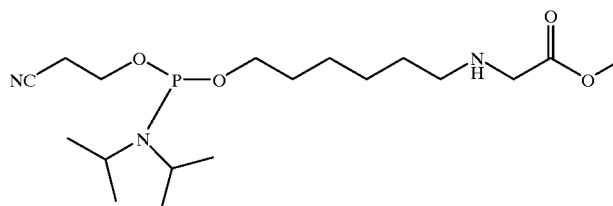
VI



VII



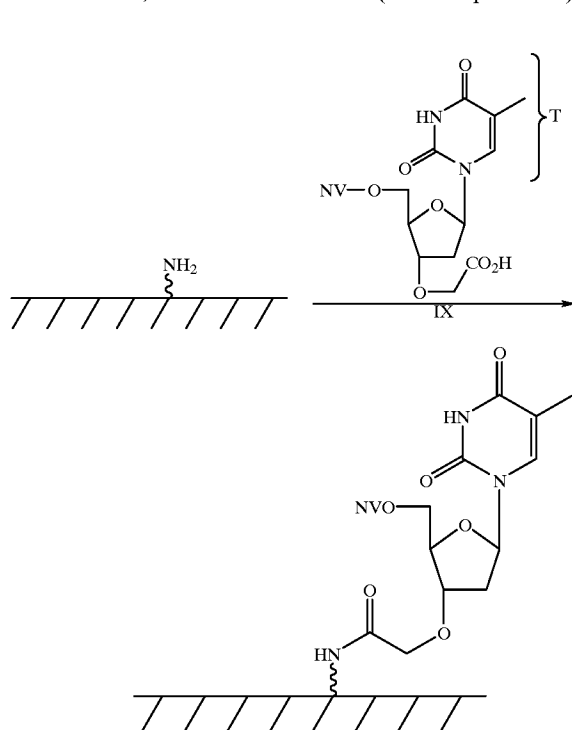
VIII



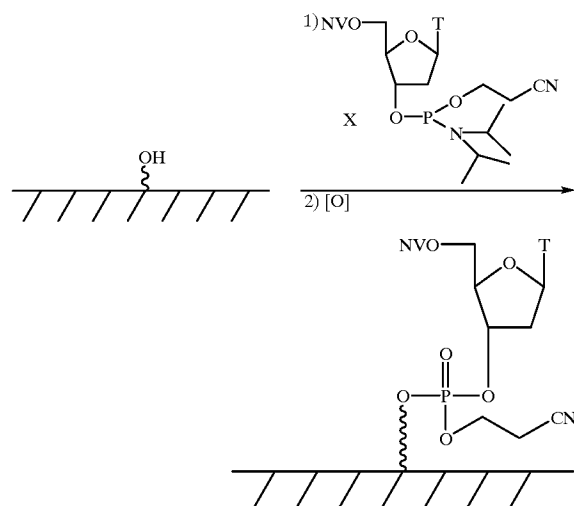
US 6,403,320 B1

81

The method of attachment of the first nucleoside to the surface of the substrate depends on the functionality of the groups at the substrate surface. If the surface is amine functionalized, an amide bond is made (see example below).



If the surface is hydroxy functionalized a phosphate bond is made (see example below)

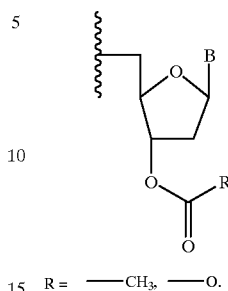


In both cases, the thymidine example is illustrated, but any one of the four phosphoramidite activated nucleosides can be used in the first step.

Photolysis of the photolabile group NV or NVOC on the 5' positions of the nucleosides is carried out at ~ 362 nm with an intensity of 14 mW/cm^2 for 10 minutes with the substrate side (side containing the photolabile group) immersed in dioxane. After the coupling of the next nucleoside is complete, the photolysis is repeated followed by another coupling until the desired oligomer is obtained.

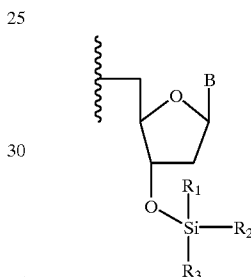
82

One of the most common 3'-O-protecting group is the ester, in particular the acetate



The groups can be removed by mild base treatment 0.1N NaOH/MeOH or $\text{K}_2\text{CO}_3/\text{H}_2\text{O/MeOH}$.

Another group used most often is the silyl ether.



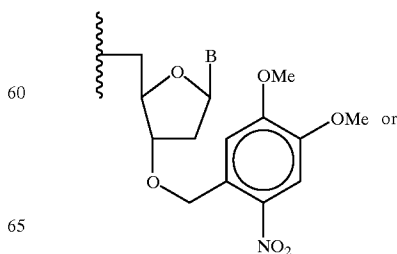
$\text{R}_1, \text{R}_2, \text{R}_3 = \text{CH}_3$

$\text{R}_1, \text{R}_2 = \text{CH}_3; \text{R}_3 = \text{---CH}_3$

$\text{R}_1, \text{R}_2, \text{R}_3 = \text{---CH}_3$

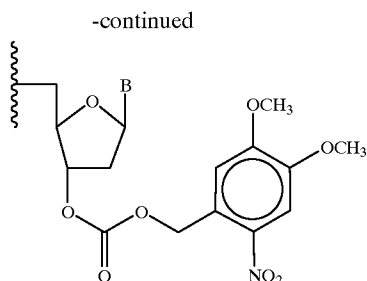
These groups can be removed by neutral conditions using 1 M tetra-*n*-butylammonium fluoride in THF or under acid conditions.

Related to photodeprotection, the nitroveratryl group could also be used to protect the 3'-position.



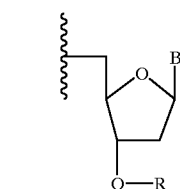
US 6,403,320 B1

83



Here, light (photolysis) would be used to remove these protecting groups.

A variety of ethers can also be used in the protection of the 3'-O-position.



R = trityl
= benzyl

Removal of these groups usually involves acid or catalytic methods.

Note that corresponding linkages and photoblocked amino acids are described in detail in U.S. Ser. No. 07/624,120, which is hereby incorporated herein by reference.

Although the specificity of interactions at particular locations will usually be homogeneous due to a homogeneous polymer being synthesized at each defined location, for certain purposes, it may be useful to have mixed polymers with a commensurate mixed collection of interactions occurring at specific defined locations, or degeneracy reducing analogues, which have been discussed above and show broad specificity in binding. Then, a positive interaction signal may result from any of a number of sequences contained therein.

As an alternative method of generating a matrix pattern on a substrate, preformed polymers may be individually attached at particular sites on the substrate. This may be performed by individually attaching reagents one at a time to specific positions on the matrix, a process which may be automated. See, e.g., U.S. Ser. No. 07/435,316 (caged biotin parent), and U.S. Ser. No. 07/612,671 (caged biotin CIP). Another way of generating a positionally defined matrix pattern on a substrate is to have individually specific reagents which interact with each specific position on the substrate. For example, oligonucleotides may be synthesized at defined locations on the substrate. Then the substrate would have on its surface a plurality of regions having homogeneous oligonucleotides attached at each position.

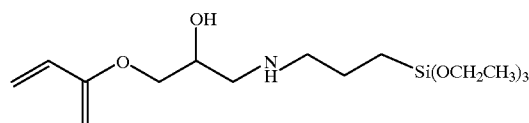
In particular, at least four different substrate preparation procedures are available for treating a substrate surface. They are the standard VLSIPS method, polymeric substrates, Durapore™, and synthetic beads or fibers. The treatment labeled "standard VLSIPS" method is described in U.S. Ser. No. 07/624,120, (automated VLSIPS), and involves applying amino-propyltriethoxysilane to a glass surface.

The polymeric substrate approach involves either of two ways of generating a polymeric substrate. The first uses a

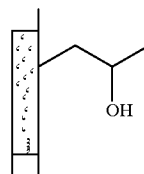
84

high concentration of aminopropyltriethoxysilane (2–20%) in an aqueous ethanol solution (95%). This allows the silane compound to polymerize both in solution and on the substrate surface, which provides a high density of amines on the surface of the glass. This density is contrasted with the standard VLSIPS method. This polymeric method allows for the deposition on the substrate surface of a monolayer due to the anhydrous method used with the aforementioned silane.

The second polymeric method involves either the coating or covalent binding of an appropriate acrylic acid polymer onto the substrate surface. In particular, e.g., in DNA synthesis, a monomer such as a hydroxypropylacrylate is used to generate a high density of hydroxyl groups on the substrate surface, allowing for the formation of phosphate bonds. An example of such a compound is shown:



The method using a Durapore™ membrane (Millipore) consists of a polyvinylidene difluoride coating with crosslinked polyhydroxylpropyl acrylate [PVDF-HPA]:



Here the building up of, e.g., a DNA oligomer, can be started immediately since phosphate bonds to the surface can be accomplished in the first step with no need for modification.

A nucleotide dimer (5'-C-T-3') has been successfully made on this substrate in our labs.

The fourth method utilizes synthetic beads or fibers. This would use another substrate, such as a teflon copolymer graft bead or fiber, which is covalently coated with an organic layer (hydrophilic) terminating in hydroxyl sites (commercially available from Molecular Brosystems, Inc.) This would offer the same advantage as the Durapore™ membrane, allowing for immediate phosphate linkages, but would give additional contour by the 3-dimensional growth of oligomers.

A matrix pattern of new reagents may be targeted to each specific oligonucleotide position by attaching a complementary oligonucleotide to which the substrate bound form is complementary. For instance, a number of regions may have homogeneous oligonucleotides synthesized at various locations. Oligonucleotide sequences complementary to each of these can be individually generated and linked to a particular specific reagents. Often these specific reagents will be antibodies. As each of these is specific for finding its complementary oligonucleotide, each of the specific reagents will bind through the oligonucleotide to the appropriate matrix position. A single step having a combination of different specific reagents being attached specifically to a particular oligonucleotide will thereby bind to its comple-

US 6,403,320 B1

85

ment at the defined matrix position. The oligonucleotides will typically then be covalently attached, using, e.g., an acridine dye, for photocrosslinking. Psoralen is a commonly used acridine dye for photocrosslinking purposes, see, e.g., Song et al. (1979) *Photochem. Photobiol.* 29:1177-1197; Cimino et al. (1985) *Ann. Rev. Biochem.* 54: 1151-1193; Parsons (1980) *Photochem. Photobiol.* 32:813-821; and Dattagupta et al. (1985) U.S. Pat. No. 4,542,102, and (1987) U.S. Pat. No. 4,713,326; each of which is hereby incorporated herein by reference. This method allows a single attachment manipulation to attach all of the specific reagents to the matrix at defined positions and results in the specific reagents being homogeneously located at defined positions. In many embodiments, the specific reagents will be antibodies.

In an alternative embodiment, antibody molecules may be used to specifically direct binding to defined positions on a substrate. The VLSIPS technology may be used to generate specific epitopes at each position on the substrate. Antibody molecules having specificity of interaction may be used to attach oligonucleotides, thereby avoiding the interference of internal polynucleotide sequences from binding to the substrate complementary oligonucleotides. In fact, the specificity of interaction for positional targeting may be achieved by use of nucleotide analogues which do not interact with the natural nucleotides. For example, other synthetic nucleotides have been made which undergo base pairing, thereby providing the specificity of targeting, but the synthetic nucleotides also do not interact with the natural biological nucleotides. Thus, synthetic oligonucleotides would be useful for attachment to biological nucleotides and specific targeting. Moreover, the VLSIPS synthetic processes would be useful in generating the VLSIPS substrate, and standard oligonucleotide synthesis could be applied, with minor modifications, to produce the complementary sequences which would be attached to other specific reagents.

D. Surface Immobilization

1. Caged Biotin

An alternative method of attaching reagents in a positionally defined matrix pattern is to use a caged biotin system. See U.S. Ser. No. 07/612,671 (caged biotin CIP), which is hereby, incorporated herein by reference, for additional details on the chemistry and application of caged biotin embodiments. In short, the caged biotin has a photosensitive blocking moiety which prevents the combination of avidin to biotin. At positions where the photo-lithographic process has removed the blocking group, high affinity biotin sites are generated. Thus, by a sequential series of photolithographic deblocking steps interspersed with exposure of those regions to appropriate biotin containing reagents, only those locations where the deblocking takes place will form an avidin-biotin interaction. Because the avidin-biotin binding is very tight, this will usually be virtually irreversible binding.

2. Crosslinked Interactions

The surface immobilization may also take place by photocrosslinking of defined oligonucleotides linked to specific reagents. After hybridization of the complementary oligonucleotides, the oligonucleotides may be crosslinked by a reagent by psoralen or another similar type of acridine dye. Other useful cross linking reagents are described in Dattagupta et al. (1985) U.S. Pat. No. 4,542,102, and (1987) U.S. Pat. No. 4,713,326.

In another embodiment, colony or phage plaque transfer of biological polymers may be transferred directly onto a silicon substrate. For example, a colony plate may be transferred onto a substrate having a generic oligonucleotide sequence which hybridizes to another generic complemen-

86

tary sequence contained on all of the vectors into which inserts are cloned. This will specifically only bind those molecules which are actually contained in the vectors containing the desired complementary sequence. This immobilization allows for producing a matrix onto which a sequence specific reagent can bind, or for other purposes. In a further embodiment, a plurality of different vectors each having a specific oligonucleotide attached to the vector may be specifically attached to particular regions on a matrix having a complementary oligonucleotide attached thereto.

VIII. Hybridization/Specific Interaction

A. General

As discussed previously in the VLSIPS parent applications, the VLSIPS substrates may be used for screening for specific interactions with sequence specific targets or probes.

In addition, the availability of substrates having the entire repertoire of possible sequences of a defined length opens up the possibility of sequencing by hybridization. This sequence may be de nova determination of an unknown sequence, particularly of nucleic acid, verification of a sequence determined by another method, or an investigation of changes in a previously sequenced gene, locating and identifying specific changes. For example, often Maxam and Gilbert sequencing techniques are applied to sequences which have been determined by Sanger and Coulson. Each of those sequencing technologies have problems with resolving particular types of sequences. Sequencing by hybridization may serve as a third and independent method for verifying other sequencing techniques. See, e.g., (1988) *Science* 242:1245.

In addition, the ability to provide a large repertoire of particular sequences allows use of short subsequence and hybridization as a means to fingerprint a sample. This may be used in a nucleic acid, as well as other polymer embodiments. For example, fingerprinting to a high degree of specificity of sequence matching may be used for identifying highly similar samples, e.g., those exhibiting high homology to the selected probes. This may provide a means for determining classifications of particular sequences. This should allow determination of whether particular genomes of bacteria, phage, or even higher cells might be related to one another.

In addition, fingerprinting may be used to identify an individual source of biological sample. See, e.g., Lander, E. (1989) *Nature*, 339:501-505, and references therein. For example, a DNA fingerprint may be used to determine whether a genetic sample arose from another individual. This would be particularly useful in various sorts of forensic tests to determine, e.g., paternity or sources of blood samples. Significant detail on the particulars of genetic fingerprinting for identification purposes are described in, e.g., Morris et al. (1989) "Biostatistical evolution of evidence from continuous allele frequency distribution DNA probes in reference to disputed paternity of identity," *J. Forensic Science* 34:1311-1317; and Neufeld et al. (1990) *Scientific American* 262:46-53; each of which is hereby incorporated herein by reference.

In another embodiment, a fingerprinting-like procedure may be used for classifying cell types by analyzing a pattern of specific nucleic acids present in the cell. A series of antibodies may be used to identify cell markers, e.g., proteins, usually on the cell surface, but intracellular markers may also be used. Antigens which are extracellularly expressed are preferred so cell lysis is unnecessary in the screening, but intracellular markers may also be useful. The markers will usually be proteins, but may be nucleic acids,

US 6,403,320 B1

87

lipids, metabolites, carbohydrates, or other cellular components. See, e.g., Winkelgren, I. (1990) *Science News* 136:234–237, which indicates extracellular DNA may be common, and suggesting that such might be characteristic of cell types, stage, or physiology. This may also be useful in defining the temporal stage of development of cells, e.g., stem cells or other cells which undergo temporal changes in development. For example, the stage of a cell, or group of cells, may be tested or defined by isolating a sample of mRNA from the population and testing to see what sequences are present in messenger populations. Direct samples, or amplified samples, may be used. Where particular mRNA or other nucleic acid sequences may be characteristic of or shown to be characteristic of particular developmental stages, physiological states, or other conditions, this fingerprinting method may define them. Similar sorts of fingerprinting may be used for determining T-cell classes or perhaps even to generate classification schemes for such proteins as major histocompatibility complex antigens. Thus, the ability to make these substrates allows both the generation of reagents which will be used for defining subclasses or classes of cells or other biological materials, but also provides the mechanisms for selecting those cells which may be found in defined population groups.

Cell classification defined by such a combination of properties, typically expression of extracellular antigens, the present invention also provides the means for isolating homogeneous population of cells. Once the antigenic determinants which define a cell class have been identified, these antigens may be used in a sequential selection process to isolate only those cells which exhibit the combination of defining structural properties.

The present invention may also be used for mapping sequences within a larger segment. This may be performed by at least two methods, particularly in reference to nucleic acids. Often, enormous segments of DNA are subcloned into a large plurality of subsequences. Ordering these subsequences may be important in determining the overlaps of sequences upon nucleotide determinations. Mapping may be performed by immobilizing particularly large segments onto a matrix using the VLSIPS technology. Alternatively, sequences may be ordered by virtue of subsequences shared by overlapping segments. See, e.g., Craig et al. (1990) *Nuc. Acids Res.* 18:2653–2660; Michiels et al. (1987) *CARIOS* 3:203–210; and Olson et al. (1986) *Proc. Natl. Acad. Sci. USA* 83:7826–7830.

B. Important Parameters

The extent of specific interaction between reagents immobilized to the VLSIPS substrate and another sequence specific reagent may be modified by the conditions of the interaction. Sequencing embodiments typically require high fidelity hybridization and the ability to discriminate perfect matching from imperfect matching. Fingerprinting and mapping embodiments may be performed using less stringent conditions, depending upon the circumstances.

For example, the specificity of antibody/antigen interaction may depend upon such parameters as pH, salt concentration, ionic composition, solvent composition, detergent composition and concentration, and chaotropic agent concentration. See, e.g., Harlow and Lane (1988) *Antibodies: A Laboratory Manual*, Cold Spring Harbor Press, New York. By careful control of these parameters, the affinity of binding may be mapped across different sequences.

In a nucleic acid hybridization embodiment, the specificity and kinetics of hybridization have been described in detail by, e.g., Wetmur and Davidson (1968) *J. Mol. Biol.*,

88

31:349–370, Britten and Kohne (1968) *Science* 161:529–530, and Kanehisa, (1984) *Nuc. Acids Res.* 12:203–213, each of which is hereby incorporated herein by reference. Parameters which are well known to affect specificity and kinetics of reaction include salt conditions, ionic composition of the solvent, hybridization temperature, length of oligonucleotide matching sequences, guanine and cytosine (GC) content, presence of hybridization accelerators, pH, specific bases found in the matching sequences, solvent conditions, and addition of organic solvents.

In particular, the salt conditions required for driving highly mismatched sequences to completion typically include a high salt concentration. The typical salt used is sodium chloride (NaCl), however, other ionic salts may be utilized, e.g., KCl. Depending on the desired stringency hybridization, the salt concentration will often be less than about 3 molar, more often less than 2.5 molar, usually less than about 2 molar, and more usually less than about 1.5 molar. For applications directed towards higher stringency matching, the salt concentrations would typically be lower. Ordinary high stringency conditions will utilize salt concentration of less than about 1 molar, more often less than about 750 millimolar, usually less than about 500 millimolar, and may be as low as about 250 or 150 millimolar.

The kinetics of hybridization and the stringency of hybridization both depend upon the temperature at which the hybridization is performed and the temperature at which the washing steps are performed. Temperatures at which steps for low stringency hybridization are desired would typically be lower temperatures, e.g., ordinarily at least about 15° C., more ordinarily at least about 20° C., usually at least about 25° C., and more usually at least about 30° C. For those applications requiring high stringency hybridization, or fidelity of hybridization and sequence matching, temperatures at which hybridization and washing steps are performed would typically be high. For example, temperatures in excess of about 35° C. would often be used, more often in excess of about 40° C., usually at least about 45° C., and occasionally even temperatures as high as about 50° C. or 60° C. or more. Of course, the hybridization of oligonucleotides may be disrupted by even higher temperatures. Thus, for stripping of targets from substrates, as discussed below, temperatures as high as 80° C., or even higher may be used.

The base composition of the specific oligonucleotides involved in hybridization affects the temperature of melting, and the stability of hybridization as discussed in the above references. However, the bias of GC rich sequences to hybridize faster and retain stability at higher temperatures can be compensated for by the inclusion in the hybridization incubation or wash steps of various buffers. Sample buffers which accomplish this result include the triethly- and trimethyl ammonium buffers. See, e.g., Wood et al. (1987) *Proc. Natl. Acad. Sci. USA*, 82:1585–1588, and Khrapko, K. et al. (1989) *FEBS Letters* 256:118–122.

The rate of hybridization can also be affected by the inclusion of particular hybridization accelerators. These hybridization accelerators include the volume exclusion agents characterized by dextran sulfate, or polyethylene glycol (PEG). Dextran sulfate is typically included at a concentration of between 1% and 40% by weight. The actual concentration selected depends upon the application, but typically a faster hybridization is desired in which the concentration is optimized for the system in question. Dextran sulfate is often included at a concentration of between 0.5% and 2% by weight or dextran sulfate at a concentration between about 0.5% and 5%. Alternatively, proteins which

US 6,403,320 B1

89

accelerate hybridization may be added, e.g., the recA protein found in *E. coli*) or other homologous proteins.

With respect to those embodiments where specific reagents are not oligonucleotides, the conditions of specific interaction would depend on the affinity of binding between the specific reagent and its target. Typically parameters which would be of particular importance would be pH, salt concentration anion and cation compositions, buffer concentration, organic solvent inclusion, detergent concentration, and inclusion of such reagents such as chaotropic agents. In particular, the affinity of binding may be tested over a variety of conditions by multiple washes and repeat scans or by using reagents with differences in binding affinity to determine which reagents bind or do not bind under the selected binding and washing conditions. The spectrum of binding affinities may provide an additional dimension of information which may be very useful in identification purposes and mapping.

Of course, the specific hybridization conditions will be selected to correspond to a discriminatory condition which provides a positive signal where desired but fails to show a positive signal at affinities where interaction is not desired. This may be determined by a number of titration steps or with a number of controls which will be run during the hybridization and/or washing steps to determine at what point the hybridization conditions have reached the stage of desired specificity.

IX. Detection Methods

Methods for detection depend upon the label selected. The criteria for selecting an appropriate label are discussed below, however, a fluorescent label is preferred because of its extreme sensitivity and simplicity. Standard labeling procedures are used to determine the positions where interactions between a sequence and a reagent take place. For example, if a target sequence is labeled and exposed to a matrix of different probes, only those locations where probes do interact with the target will exhibit any signal. Alternatively, other methods may be used to scan the matrix to determine where interaction takes place. Of course, the spectrum of interactions may be determined in a temporal manner by repeated scans of interactions which occur at each of a multiplicity of conditions. However, instead of testing each individual interaction separately, a multiplicity of sequence interactions may be simultaneously determined on a matrix.

A. Labeling Techniques

The target polynucleotide may be labeled by any of a number of convenient detectable markers. A fluorescent label is preferred because it provides a very strong signal with low background. It is also optically detectable at high resolution and sensitivity through a quick scanning procedure. Other potential labeling moieties include, radioisotopes, chemiluminescent compounds, labeled binding proteins, heavy metal atoms, spectroscopic markers, magnetic labels, and linked enzymes.

Another method for labeling may bypass any label of the target sequence. The target may be exposed to the probes, and a double strand hybrid is formed at those positions only. Addition of a double strand specific reagent will detect where hybridization takes place. An intercalative dye such as ethidium bromide may be used as long as the probes themselves do not fold back on themselves to a significant extent forming hairpin loops. See, e.g., Sheldon et al. (1986) U.S. Pat. No. 4,582,789. However, the length of the hairpin loops in short oligonucleotide probes would typically be insufficient to form a stable duplex.

90

In another embodiment, different targets may be simultaneously sequenced where each target has a different label. For instance, one target could have a green fluorescent label and a second target could have a red fluorescent label. The scanning step will distinguish sites of binding of the red label from those binding the green fluorescent label. Each sequence can be analyzed independently from one another.

Suitable chromogens will include molecules and compounds which absorb light in a distinctive range of wavelengths so that a color may be observed, or emit light when irradiated with radiation of a particular wave length or wave length range, e.g., fluorescers. Biliproteins, e.g., ficeerythrin, may also serve as labels.

A wide variety of suitable dyes are available, being primary chosen to provide an intense color with minimal absorption by their surroundings. Illustrative dye types include quinoline dyes, triarylmethane dyes, acridine dyes, alizarine dyes, phthaleins, insect dyes, azo dyes, anthraquinoid dyes, cyanine dyes, phenazathionium dyes, and phenazonium dyes.

A wide variety of fluorescers may be employed either by themselves or in conjunction with quencher molecules. Fluorescers of interest fall into a variety of categories having certain primary functionalities. These primary functionalities include 1- and 2-aminonaphthalene, p,p'-diaminostilbenes, pyrenes, quaternary phenanthridine salts, 9-aminoacridines, p,p'-diaminobenzophenone imines, anthracenes, oxacarbocyanine, merocyanine, 3-aminoequilenin, perylene, bis-benzoxazole, bis-p-oxazolyl benzene, 1,2-benzophenazin, retinol, bis-3-aminopyridinium salts, hellebrigenin, tetracycline, sterophenol, benzimidazoalylphenylamine, 2-oxo-3-chromen, indole, xanthen, 7-hydroxycoumarin, phenoxazine, salicylate, strophanthidin, porphyrins, triarylmethanes and flavin. Individual fluorescent compounds which have functionalities for linking or which can be modified to incorporate such functionalities include, e.g., dansyl chloride; fluoresceins such as 3,6-dihydroxy-9-phenylxanthhydrol; rhodamineisothiocyanate; N-phenyl 1-amino-8-sulfonatophthalene; N-phenyl 2-amino-6-sulfonatophthalene; 4-acetamido-4-isothiocyanatostilbene-2,2'-disulfonic acid; pyrene-3-sulfonic acid; 2-toluidinonaphthalene-6-sulfonate; N-phenyl, N-methyl 2-aminoaphthalene-6-sulfonate; ethidium bromide; stebrine; auroamine-0,2-(9'-anthroyl)palmitate; dansyl phosphatidylethanolamine; N,N'-dioctadecyl oxacarbocyanine; N,N'-dihexyl oxacarbocyanine; merocyanine, 4-(3'pyrenyl) butyrate; d-3-aminodesoxy-equilenin; 12-(9'-anthroyl) stearate; 2-methylanthracene; 9-vinylanthracene; 2,2'-(vinylene-p-phenylene)bisbenzoxazole; p-bis[2-(4-methyl-5-phenyl-oxazolyl)]-benzene; 6-dimethylamino-1,2-benzophenazin; retinol; bis(3'-aminopyridinium) 1,10-decandiyl diiodide; sulfonaphthylhydrazine of hellebrigenin; chlorotetracycline; N-(7-dimethylamino-4-methyl-2-oxo-3-chromenyl)maleimide; N-[p-(2-benzimidazolyl)-phenyl] maleimide; N-(4-fluoranthyl)maleimide; bis(homovanillic acid); resazarin; 4-chloro-7-nitro-2,1,3-benzooxadiazole; merocyanine 540; resorufin; rose bengal; and 2,4-diphenyl-3(2H)-furanone.

Desirably, fluorescers should absorb light above about 300 nm, preferably about 350 nm, and more preferably above about 400 nm, usually emitting at wavelengths greater than about 10 nm higher than the wavelength of the light absorbed. It should be noted that the absorption and emission characteristics of the bound dye may differ from the unbound dye. Therefore, when referring to the various wavelength ranges and characteristics of the dyes, it is

US 6,403,320 B1

91

intended to indicate the dyes as employed and not the dye which is unconjugated and characterized in an arbitrary solvent.

Fluorescers are generally preferred because by irradiating a fluorescer with light, one can obtain a plurality of emissions. Thus, a single label can provide for a plurality of measurable events.

Detectable signal may also be provided by chemiluminescent and bioluminescent sources. Chemiluminescent sources include a compound which becomes electronically excited by a chemical reaction and may then emit light which serves as the detectible signal or donates energy to a fluorescent acceptor. A diverse number of families of compounds have been found to provide chemiluminescence under a variety of conditions. One family of compounds is 2,3-dihydro-1,4-phthalazinedione. The most popular compound is luminol, which is the 5-amino compound. Other members of the family include the 5-amino-6,7,8-trimethoxy- and the dimethylamino[ca]benz analog. These compounds can be made to luminesce with alkaline hydrogen peroxide or calcium hypochlorite and base. Another family of compounds is the 2,4,5-triphenylimidazoles, with lophine as the common name for the parent product. Chemiluminescent analogs include para-dimethylamino and -methoxy substituents. Chemiluminescence may also be obtained with oxalates, usually oxalyl active esters, e.g., p-nitrophenyl and a peroxide, e.g., hydrogen peroxide, under basic conditions. Alternatively, luciferins may be used in conjunction with luciferase or lucigenins to provide bioluminescence.

Spin labels are provided by reporter molecules with an unpaired electron spin which can be detected by electron spin resonance (ESR) spectroscopy. Exemplary spin labels include organic free radicals, transitional metal complexes, particularly vanadium, copper, iron, and manganese, and the like. Exemplary spin labels include nitroxide free radicals.

B. Scanning System

With the automated detection apparatus, the correlation of specific positional labeling is converted to the presence on the target of sequences for which the reagents have specificity of interaction. Thus, the positional information is directly converted to a database indicating what sequence interactions have occurred. For example, in a nucleic acid hybridization application, the sequences which have interacted between the substrate matrix and the target molecule can be directly listed from the positional information. The detection system used is described in U.S. Ser. No. 07/649,642 (VLSIPS CIP); and U.S. Ser. No. 07/624,120, (automated VLSIPS). Although the detection described therein is a fluorescence detector, the detector may be replaced by a spectroscopic or other detector. The scanning system may make use of a moving detector relative to a fixed substrate, a fixed detector with a moving substrate, or a combination. Alternatively, mirrors or other apparatus can be used to transfer the signal directly to the detector. See, e.g., U.S. Ser. No. 07/624,120, (VLSIPS), which is hereby incorporated herein by reference.

The detection method will typically also incorporate some signal processing to determine whether the signal at a particular matrix position is a true positive or may be a spurious signal. For example, a signal from a region which has actual positive signal may tend to spread over and provide a positive signal in an adjacent region which actually should not have one. This may occur, e.g., where the scanning system is not properly discriminating with sufficiently high resolution in its pixel density to separate the two regions. Thus, the signal over the spatial region may be

92

evaluated pixel by pixel to determine the locations and the actual extent of positive signal. A true positive signal should, in theory, show a uniform signal at each pixel location. Thus, processing by plotting number of pixels with actual signal intensity should have a clearly uniform signal intensity. Regions where the signal intensities show a fairly wide dispersion, may be particularly suspect and the scanning system may be programmed to more carefully scan those positions.

In another embodiment, as the sequence of a target is determined at a particular location, the overlap for the sequence would necessarily have a known sequence. Thus, the system can compare the possibilities for the next adjacent position and look at these in comparison with each other. Typically, only one of the possible adjacent sequences should give a positive signal and the system might be programmed to compare each of these possibilities and select that one which gives a strong positive. In this way, the system can also simultaneously provide some means of measuring the reliability of the determination by indicating what the average signal to background ratio actually is.

More sophisticated signal processing techniques can be applied to the initial determination of whether a positive signal exists or not. See, e.g., U.S. Ser. No. 07/624,120, (automated VLSIPS).

From a listing of those sequences which interact, data analysis may be performed on a series of sequences. For example, in a nucleic acid sequence application, each of the sequences may be analyzed for their overlap regions and the original target sequence may be reconstructed from the collection of specific subsequences obtained therein. Other sorts of analyses for different applications may also be performed, and because the scanning system directly interfaces with a computer the information need not be transferred manually. This provides for the ability to handle large amounts of data with very little human intervention. This, of course, provides significant advantages over manual manipulations. Increased throughput and reproducibility is thereby provided by the automation of vast majority of steps in any of these applications.

XI. Data Analysis

A. General

Data analysis will typically involve aligning the proper sequences with their overlaps to determine the target sequence. Although the target "sequence" may not specifically correspond to any specific molecule, especially where the target sequence is broken and fragmented up in the sequencing process, the sequence corresponds to a contiguous sequence of the subfragments.

The data analysis can be performed by a computer using an appropriate program. See, e.g., Drmanac, R. et al. (1989) *Genomics* 4:114-128; and a commercially available analysis program available from the Genetic Engineering Center, P.O. Box 794, 11000 Belgrade, Yugoslavia. Although the specific manipulations necessary to reassemble the target sequence from fragments may take many forms, one embodiment uses a sorting program to sort all of the subsequences using a defined hierarchy. The hierarchy need not necessarily correspond to any physical hierarchy, but provides a means to determine, in order, which subfragments have actually been found in the target sequence. In this manner, overlaps can be checked and found directly rather than having to search throughout the entire set after each selection process. For example, where the oligonucleotide probes are 10-mers, the first 9 positions can be sorted. A particular subsequence can be selected as in the examples, to determine where the process starts. As analogous to the

US 6,403,320 B1

93

theoretical example provided above, the sorting procedure provides the ability to immediately find the position of the subsequence which contains the first 9 positions and can compare whether there exists more than 1 subsequence during the first 9 positions. In fact, the computer can easily generate all of the possible target sequences which contain given combination of subsequences. Typically there will be only one, but in various situations, there will be more.

An exemplary flow chart for a sequencing program is provided in FIG. 4. In general terms, the program provides for automated scanning of the substrate to determine the positions of probe and target interaction. Simple processing of the intensity of the signal may be incorporated to filter out clearly spurious signals. The positions with positive interaction are correlated with the sequence specificity of specific matrix positions, to generate the set of matching subsequences. This information is further correlated with other target sequence information, e.g., restriction fragment analysis. The sequences are then aligned using overlap data, thereby leading to possible corresponding target sequences which will, optimally, correspond to a single target sequence.

B. Hardware

A variety of computer systems may be used to run a sequencing program. The program may be written to provide both the detecting and scanning steps together and will typically be dedicated to a particular scanning apparatus. However, the components and functional steps may be separated and the scanning system may provide an output, e.g., through tape or an electronic connection into a separate computer which separately runs the sequencing analysis program. The computer may be any of a number of machines provided by standard computer manufacturers, e.g., IBM compatible machines, Apple™ machines, VAX machines, and others, which may often use a UNIX™ operating system. Of course, the hardware used to run the analysis program will typically determine what programming language would be used.

C. Software

Software would be easily developed by a person of ordinary skill in the programming art, following the flow chart provided, or based upon the input provided and the desired result.

Of course, an exemplary embodiment is a polynucleotide sequence system. However, the theoretical and mathematical manipulations necessary for data analysis of other linear molecules, such as polypeptides, carbohydrates, and various other polymers are conceptually similar. Simple branching polymers will usually also be sequencable using similar technology. However, where there is branching, it may be desired that additional recognition reagents be used to determine the nature and location of branches. This can easily be provided by use of appropriate specific reagents which would be generated by methods similar to those used to produce specific reagents for linear polymers.

XII. Substrate Reuse

Where a substrate is made with specific reagents that are relatively insensitive to the handling and processing steps involved in a single cycle of use, the substrate may often be reused. The target molecules are usually stripped off of the solid phase specific recognition molecules. Of course, it is preferred that the manipulations and conditions be selected as to be mild and to not affect the substrate. For example, if a substrate is acid labile, a neutral pH would be preferred in all handling steps. Similar sensitivities would be carefully respected where recycling is desired.

94

A. Removal of Label

Typically for a recycling, the previously attached specific interaction would be disrupted and removed. This will typically involve exposing the substrate to conditions under which the interaction between probe and target is disrupted. Alternatively, it may be exposed to conditions where the target is destroyed. For example, where the probes are oligonucleotides and the target is a polynucleotide, a heating and low salt wash will often be sufficient to disrupt the interactions. Additional reagents may be added such as detergents and organic or inorganic solvents which disrupt the interaction between the specific reagents and target. In an embodiment where the specific reagents are antibodies, the substrate may be exposed to a gentle detergent which will denature the specific binding between the antibody and its target. The conditions are selected to avoid severe disruption or destruction of the structure of the antibody and to maintain the specificity of the antibody binding site. Conditions with specific pH, detergent concentration, salt concentration, ionic concentration, and other parameters may be selected which disrupt the specific interactions.

B. Storage and Preservation

As indicated above, the matrix will typically be maintained under conditions where the matrix itself and the linkages and specific reagents are preserved. Various specific preservatives may be added which prevent degradation. For example, if the reagents are acid or base labile, a neutral pH buffer will typically be added. It is also desired to avoid destruction of the matrix by growth of organisms which may destroy organic reagents attached thereto. For this reason, a preservative such as cyanide or azide may be added. However, the chemical preservative should also be selected to preserve the chemical nature of the linkages and other components of the substrate. Typically, a detergent may also be included.

C. Processes to Avoid Degradation of Oligomers

In particular, a substrate comprising a large number of oligomers will be treated in a fashion which is known to maintain the quality and integrity of oligonucleotides. These include storing the substrate in a carefully controlled environment under conditions of lower temperature, cation depletion (EDTA and EGTA), sterile conditions, and inert argon or nitrogen atmosphere.

XIII. Integrated Sequencing Strategy

A. Initial Mapping Strategy

As indicated above, although the VLSIPS may be applied to sequencing embodiments, it is often useful to integrate other concepts to simplify the sequencing. For example, nucleic acids may be easily sequenced by careful selection of the vectors and hosts used for amplifying and generating the specific target sequences. For example, it may be desired to use specific vectors which have been designed to interact most efficiently with the VLSIPS substrate. This is also important in fingerprinting and mapping strategies. For example, vectors may be carefully selected having particular complementary sequences which are designed to attach to a genetic or specific oligomer on the substrate. This is also applicable to situations where it is desired to target particular sequences to specific locations on the matrix.

In one embodiment, unnatural oligomers may be used to target natural probes to specific locations on the VLSIPS substrate. In addition, particular probes may be generated for the mapping embodiment which are designed to have specific combinations of characteristics. For example, the construction of a mapping substrate may depend upon use of another automated apparatus which takes clones isolated from a chromosome walk and attaches them individually or in bulk to the VLSIPS substrate.

US 6,403,320 B1

95

In another embodiment, a variety of specific vectors having known and particular "targeting" sequences adjacent the cloning sites may be individually used to clone a selected probe, and the isolated probe will then be targetable to a site on the VLSIPS substrate with a sequence complementary to the "target" sequence.

B. Selection of Smaller Clones

In the fingerprinting and mapping embodiments, the selection of probes may be very important. Significant mathematical analysis may be applied to determine which specific sequences should be used as those probes. Of course, for fingerprinting use, these sequences would be most desired that show significant heterogeneity across the human population. Selection of the specific sequences which would most favorably be utilized will tend to be single copy sequences within the genome.

Various hybridization selection procedures may be applied to select sequences which tend not to be repeated within a genome, and thus would tend to be conserved across individuals. For example, hybridization selections may be made for non-repetitive and single copy sequences. See, e.g., Britten and Kohne (1968) "Repeated Sequences in DNA," *Science* 161:529-540. On the other hand, it may be desired under certain circumstances to use repeated sequences. For example, where a fingerprint may be used to identify or distinguish different species, or where repetitive sequences may be diagnostic of specific species, repetitive sequences may be desired for inclusion in the fingerprinting probes. In either case, the sequencing capability will greatly assist in the selection of appropriate sequences to be used as probes.

Also as indicated above, various means for constructing an appropriate substrate may involve either mechanical or automated procedures. The standard VLSIPS automated procedure involves synthesizing oligonucleotides or short polymers directly on the substrate. In various other embodiments, it is possible to attach separately synthesized reagents onto the matrix in an ordered array. Other circumstances may lend themselves to transfer a pattern from a petri plate onto a solid substrate. Also, there are methods for site specifically directing collections of reagents to specific locations using unnatural nucleotides or equivalent sorts of targeting molecules.

While a brute force manual transfer process may be utilized sequentially attaching various samples to successive positions, instrumentation for automating such procedures may also be devised. The automated system for performing such would preferably be relatively easily designed and conceptually easily understood.

XIV. Commercial Applications

A. Sequencing

As indicated above, sequencing may be performed either de novo or as a verification of another sequencing method. The present hybridization technology provides the ability to sequence nucleic acids and polynucleotides de novo, or as a means to verify either the Maxam and Gilbert chemical sequencing technique or Sanger and Coulson dideoxy-sequencing techniques. The hybridization method is useful to verify sequencing determined by any other sequencing technique and to closely compare two similar sequences, e.g., to identify and locate sequence differences.

Besides polynucleotide sequencing, the present invention also provides means for sequencing other polymers. This includes polypeptides, carbohydrates, synthetic organic polymers, and other polymers. Again, the sequencing may be either verification or de novo.

Of course, sequencing of can be very important in many different sorts of environments. For example, it will be

96

useful in determining the genetic sequence of particular markers in various individuals. In addition, polymers may be used as markers or for information containing molecules to encode information. For example, a short polynucleotide sequence may be included in large bulk production samples indicating the manufacturer, date, and location of manufacture of a product. For example, various drugs may be encoded with this information with a small number of molecules in a batch. For example, a pill may have somewhere from 10 to 100 to 1,000 or more very short and small molecules encoding this information. When necessary, this information may be decoded from a sample of the material using a polymerase chain reaction (PCR) or other amplification method. This encoding system may be used to provide the origin of large bulky samples without significantly affecting the properties of those samples. For example, chemical samples may also be encoded by this method thereby providing means for identifying the source and manufacturing details of lots. The origin of bulk hydrocarbon samples may be encoded. Production lots of organic compounds such as benzene or plastics may be encoded with a short molecule polymer. Food stuffs may also be encoded using similar marking molecules. Even toxic waste samples can be encoded determining the source or origin. In this way, proper disposal can be traced or more easily enforced.

Similar sorts of encoding may be provided by fingerprinting-type analysis. Whether the resolution is absolute or less so, the concept of coding information on molecules such as nucleic acids, which can be amplified and later decoded, may be a very useful and important application.

This technology also provides the ability to include markers for origins of biological materials. For example, a patented animal line may be transformed with a particular unnatural sequence which can be traced back to its origin. With a selection of multiple markers, the likelihood could be negligible that a combination of markers would have independently arisen from a source other than the patented or specifically protected source. This technique may provide a means for tracing the actual origin of particular biological materials. Bacteria, plants, and animals will be subject to marking by such encoding sequences.

B. Fingerprinting

As indicated above, fingerprinting technology may also be used for data encryption. Moreover, fingerprinting allows for significant identification of particular individuals. Where the fingerprinting technology is standardized, and used for identification of large numbers of people, related equipment and peripheral processing will be developed to accompany the underlying technology. For example, specific equipment may be developed for automatically taking a biological sample and generating or amplifying the information molecules within the sample to be used in fingerprinting analysis. Moreover, the fingerprinting substrate may be mass produced using particular types of automatic equipment. Synthetic equipment may produce the entire matrix simultaneously by stepwise synthetic methods as provided by the VLSIPS technology. The attachment of specific probes onto a substrate may also be automated, e.g., making use of the caged biotin technology. See, e.g., U.S. Ser. No. 07/612,671 (caged biotin CIP). As indicated above, there are automated methods for actually generating the matrix and substrate with distinct sequence reagents positionally located at each of the matrix positions. Where such reagents are, e.g., unnatural amino acids, a targeting function may be utilized which does not interfere with aa natural nucleotide functionality.

US 6,403,320 B1

97

In addition, peripheral processing may be important and may be dedicated to this specific application. Thus, automated equipment for producing the substrates may be designed, or particular systems which take in a biological sample and output either a computer readout or an encoded instrument, e.g., a card or document which indicates the information and can provide that information to others. An identification having a short magnetic strip with a few million bits may be used to provide individual identification and important medical information useful in a medical emergency.

In fact, data banks may be set up to correlate all of this information of fingerprinting with medical information. This may allow for the determination of correlations between various medical problems and specific DNA sequences. By collating large populations of medical records with genetic information, genetic propensities and genetic susceptibilities to particular medical conditions may be developed. Moreover, with standardization of substrates, the micro encoding data may be also standardized to reproduce the information from a centralized data bank or on an encoding device carried on an individual person. On the other hand, if the fingerprinting procedure is sufficiently quick and routine, every hospital may routinely perform a fingerprinting operation and from that determine many important medical parameters for an individual.

In particular industries, the VLSIPS sequencing, fingerprinting, or mapping technology will be particularly appropriate. As mentioned above, agricultural livestock suppliers may be able to encode and determine whether their particular strains are being used by others. By incorporating particular markers into their genetic stocks, the markers will indicate origin of genetic material. This is applicable to seed producers, livestock producers, and other suppliers of medical or agricultural biological materials.

This may also be useful in identifying individual animals or plants. For example, these markers may be useful in determining whether certain fish return to their original breeding grounds, whether sea turtles always return to their original birthplaces, or to determine the migration patterns and viability of populations of particular endangered species. It would also provide means for tracking the sources of particular animal products. For example, it might be useful for determining the origins of controlled animal substances such as elephant ivory or particular bird populations whose importation or exportation is controlled.

As indicated above, polymers may be used to encode important information on source and batch and supplier. This is described in greater detail, e.g., "Applications of PCR to industrial problems," (1990) in *Chemical and Engineering News* 68:145, which is hereby incorporated herein by reference. In fact, the synthetic method can be applied to the storage of enormous amounts of information. Small substrates may encode enormous amounts of information, and its recovery will make use of the inherent replication capacity. For example, on regions of $10\ \mu\text{m} \times 10\ \mu\text{m}$, $1\ \text{cm}^2$ has 10^6 regions. An theory, the entire human genome could be attached in 1000 nucleotide segments on a $3\ \text{cm}^2$ surface. Genomes of endangered species may be stored on these substrates.

Fingerprinting may also be used for genetic tracing or for identifying individuals for forensic science purposes. See, e.g., Morris, J. et al. (1989) "Biostatistical Evaluation of Evidence From Continuous Allele Frequency Distribution DNA Probes in Reference to Disputed Paternity and Identity," *J. Forensic Science* 34:1311-1317, and references provided therein; each of which is hereby incorporated herein by reference.

98

In addition, the high resolution fingerprinting allows the distinguishability to high resolution of particular samples. As indicated above, new cell classifications may be defined based on combinations of a large number of properties. Similar applications will be found in distinguishing different species of animals or plants. In fact, microbial identification may become dependent on characterization of the genetic content. Tumors or other cells exhibiting abnormal physiology will be detectable by use of the present invention. Also, knowing the genetic fingerprint of a microorganism may provide very useful information on how to treat an infection by such organism.

Modifications of the fingerprint embodiments may be used to diagnose the condition of the organism. For example, a blood sample is presently used for diagnosing any of a number of different physiological conditions. A multi-dimensional fingerprinting method made available by the present invention could become a routine means for diagnosing an enormous number of physiological features simultaneously. This may revolutionize the practice of medicine in providing information on an enormous number of parameters together at one time. In another way, the genetic predisposition may also revolutionize the practice of medicine providing a physician with the ability to predict the likelihood of particular medical conditions arising at any particular moment. It also provides the ability to apply preventative medicine.

The present invention might also find application in use for screening new drugs and new reagents which may be very important in medical diagnosis or other applications. For example, a description of generating a population of monoclonal antibodies with defined specificities may be very useful for producing various drugs or diagnostic reagents.

Also available are kits with the reagents useful for performing sequencing, fingerprinting, and mapping procedures. The kits will have various compartments with the desired necessary reagents, e.g., substrate, labeling reagents for target samples, buffers, and other useful accompanying products.

C. Mapping

The present invention also provides the means for mapping sequences within enormous stretches of sequence. For example, nucleotide sequences may be mapped within enormous chromosome size sequence maps. For example, it would be possible to map a chromosomal location within the chromosome which contains hundreds of millions of nucleotide base pairs. In addition, the mapping and fingerprinting embodiments allow for testing of chromosomal translocations, one of the standard problems for which amniocentesis is performed.

Thus, the present invention provides a powerful tool and the means for performing sequencing, fingerprinting, and mapping functions on polymers. Although most easily and directly applicable to polynucleotides, polypeptides, carbohydrates, and other sorts of molecules can be advantageously utilized using the present technology.

The present invention will be better understood by reference to the following illustrative examples. The following examples are offered by way of illustration and not by way of limitation.

EXPERIMENTAL

I. Sequencing

- A. polynucleotide
- B. polypeptide
- C. short peptide

1. Herz antibody identification

US 6,403,320 B1

99

II. Fingerprinting

- A. polynucleotide fingerprint
- B. peptide fingerprint
- C. cell classification scheme
- D. temporal development scheme
 - 1. developmental antigens
 - 2. developmental mRNA expression
- E. diagnostic test
 - 1. viral identification
 - 2. bacterial identification
 - 3. other microbiological identifications
 - 4. allergy test (immobilized antigens)
- F. individual (animal/plant) identification
 - 1. genetic
 - 2. immunological
- G. genetic screen
 - 1. test alleles with markers
 - 2. amniocentesis

III. Mapping

- A. positionally located clones (caged biotin)
 - 1. short probes, long targets
 - 2. long targets, short probes
- B. positionally defined clones

IV. Conclusion

Relevant applications whose techniques are incorporated herein by reference are Pirrung, et al., U.S. Ser. No. 07/362,901 (VLSIPS parent), filed Jun. 7, 1989; Pirrung et al, U.S. Ser. No. 07/492,462 (VLSIPS CIP), filed Mar. 7, 1990; Barrett, et al., U.S. Ser. No. 07/435,316 (caged biotin) filed Nov. 13, 1989; Barrett, et al., U.S. Ser. No. 07/612,671 (caged biotin CIP), filed Nov. 13, 1990; and commonly assigned and simultaneously filed applications U.S. Ser. No. 07/624,120, (automated VLSIPS) and U.S. Ser. No. 07/626,730, (sequencing by synthesis).

Also, additional relevant techniques are described, e.g., in Sambrook, J., et al. (1989) *Molecular Cloning: a Laboratory Manual*, 2d Ed., vols 1-3, Cold Spring Harbor Press, New York; Greenstein and Winitz (196) *Chemistry of the Amino Acids*, Wiley and Sons, New York; Bodzansky, M. (1988) *Peptide Chemistry: a Practical Textbook*, Springer-Verlag, New York; Harlow and Lane (1988) *Antibodies: A Labora-*

100

- tory Manual*, Cold Spring Harbor Press, New York; Glover, D. (ed.) (1987) *DNA Cloning: A Practical Approach*, vols 1-3, IRL Press, Oxford; Bishop and Rawlings (1987) *Nucleic Acid and Protein Sequence Analysis: A Practical Approach*, IRL Press, Oxford; Hames and Higgins (1985) *Nucleic Acid Hybridisation: A Practical Approach*, IRL Press, Oxford; Wu et al. (1989) *Recombinant DNA Methodology*, Academic Press, San Diego; Goding (1986) *Monoclonal Antibodies: Principles and Practice*, (2ed.), Academic Press, San Diego; Finegold and Barron (1986) *Bailey and Scott's Diagnostic Microbiology*, (7th ed.), Mosby Co., St. Louis; Collins et al. (1989) *Microbiological Methods*, (6th ed.), Butterworth, London; Chaplin and Kennedy (1986) *Carbohydrate Analysis: A Practical Approach*, IRL Press, Oxford; Van Dyke (ed.) (1985) *Bioluminescence and Chemiluminescence: Instruments and Applications*, vol 1, CRC Press, Boca Rotan; and Ausubel et al. (ed.) (1990) *Current Protocols in Molecular Biology*, Greene Publishing and Wiley-Interscience, New York; each of which is hereby incorporated herein by reference.

The following examples are provided to illustrate the efficacy of the inventions herein. All operations were conducted at about ambient temperatures and pressures unless indicated to the contrary.

I. Sequencing

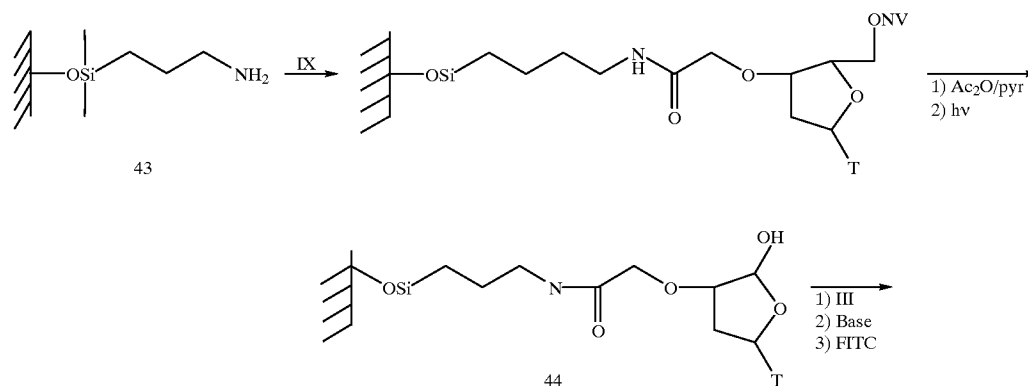
A. Polynucleotide

1. HPLC of the Photolysis of 5'-O-nitroveratryl-thymidine.

In order to determine the time for photolysis of 5'-O-nitroveratryl thymidine to thymidine a 100 μ M solution of NV-Thym-OH (5'-O-nitroveratryl thymidine) in dioxane was made and ~200 μ l aliquots were irradiated (in a quartz cuvette 1 cm \times 2 mm) at 362.3 nm for 20 sec, 40 sec, 60 sec, 2 min, 5 min, 10 min, 15 min, and 20 min. The resulting irradiated mixtures were then analyzed by HPLC using a Varian MicroPak SP column (C₁₈ analytical) at a flow rate of 1 ml/min and a solvent system of 40% CH₃CN and 60% water. Thymidine has a retention time of 1.2 min and NVO-Thym-OH has a retention time of 2.1 min. It was seen that after 10 min of exposure the deprotection was complete.

2. Preparation and Detection of Thymidine-Cytidine Dimer (FITC)

The reaction is illustrated:

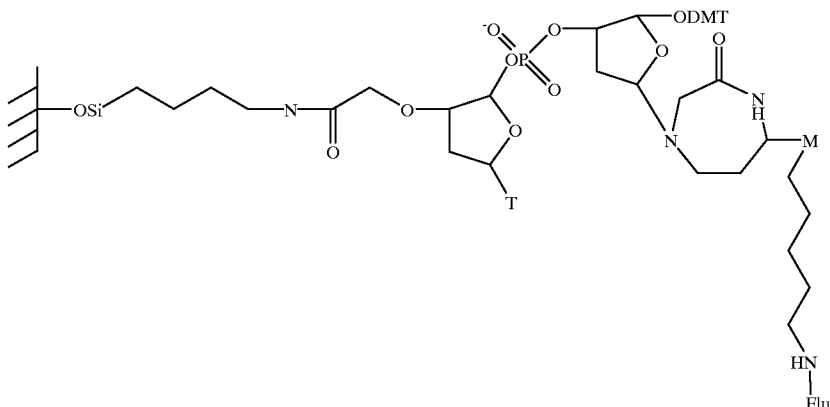


US 6,403,320 B1

101

102

-continued



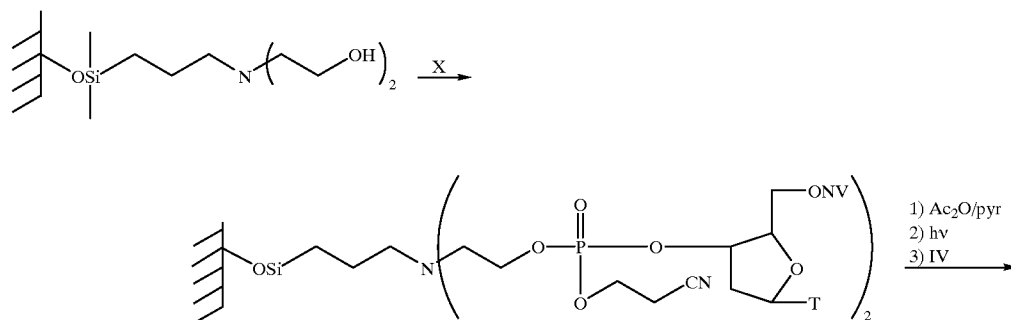
To an aminopropylated glass slide (standard VLSIPS) was added a mixture of the following:

12.2 mg of NVO-Thym-CO₂H (IX)
 3.4 mg of HOBt (N-hydroxybenztriazol)
 8.8 μ l DIEA (Diisopropylethylamine)
 11.1 mg BOP reagent
 2.5 ml DMF

After 2 h coupling time (standard VLSIPS) the plate was washed, acetylated with acetic anhydride/pyridine, washed, dried, and photolyzed in dioxane at 362 nm at 14 mW/cm² for 10 min using a 500 μ m checkerboard mask. The slide was then taken and treated with a mixture of the following:

107 mg of Fmoc-amine modified C (III)
 21 mg of tetrazole
 1 ml anhydrous CH₃CN

After being treated for approximately 8 min, the slide was washed off with CH₃CN, dried, and oxidized with I₂/H₂O/THF/lutidine for 1 min. The slide was again washed, dried, and treated for 30 min with a 20% solution of DBU in DMF. After thorough rinsing of the slide, it was next exposed to a FITC solution (1 mM fluorescein isothiocyanate [FITC] in DMF) for 50 min, then washed, dried, and examined by fluorescence microscopy. This reaction is illustrated:

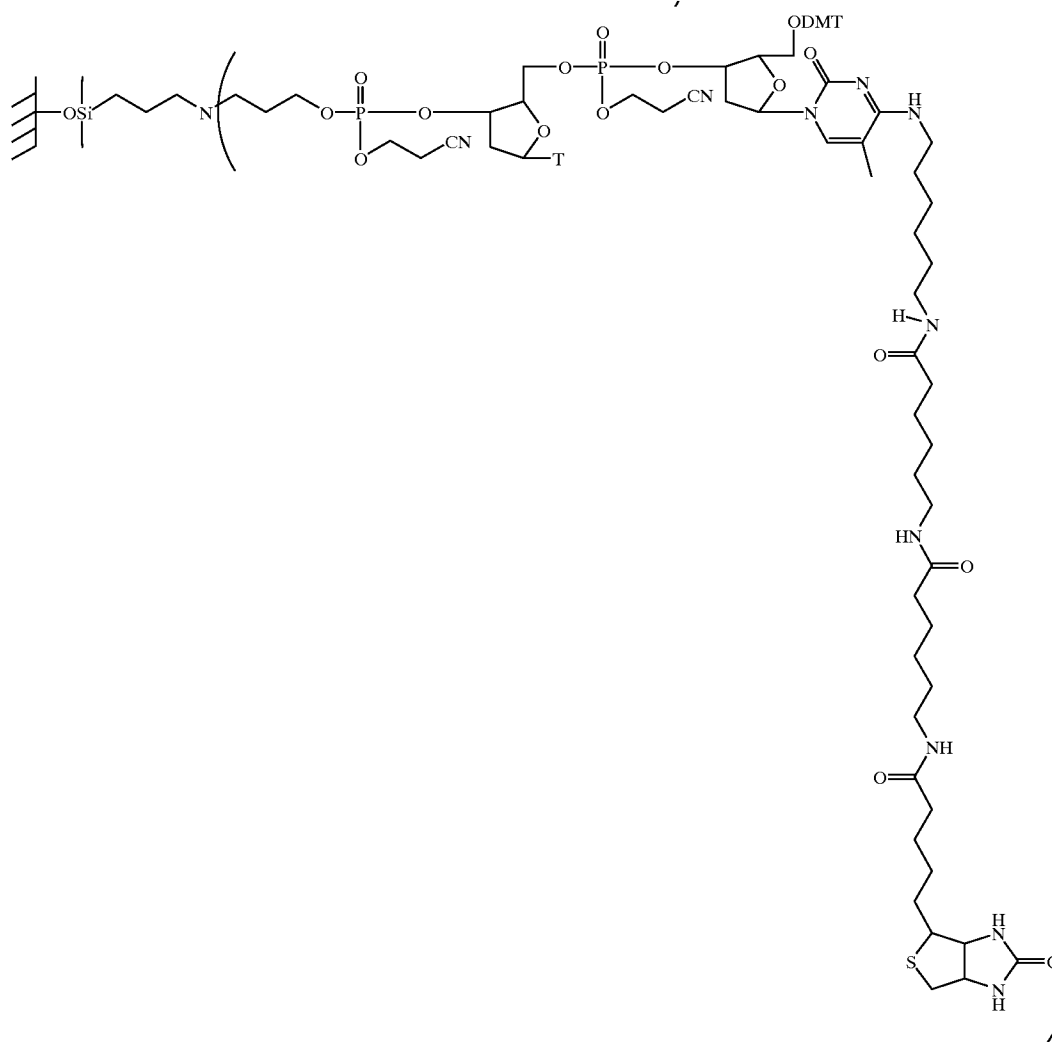


US 6,403,320 B1

103

104

-continued



3. Preparation and Detection of Thymidine-Cytidine Dimer (Biotin)

An aminopropyl glass slide, was soaked in a solution of ethylene oxide (20% in DMF) to generate a hydroxylated surface. The slide was added a mixture of the following:

32 mg of NVO-T-OCED (X)

11 mg of tetrazole

0.5 ml of anhydrous CH_3CN

After 8 min the plate was then rinsed with acetonitrile, then oxidized with $\text{I}_2/\text{H}_2\text{O}/\text{THF}/\text{lutidine}$ for 1 min, washed and dried. The slide was then exposed to a 1:3 mixture of acetic anhydride:pyridine for 1 h, then washed and dried. The substrate was then photolyzed in dioxane at 362 nm at 14 mW/cm² for 10 min using a 500 μm checkerboard mask, dried, and then treated with a mixture of the following:

65 mg of biotin modified C (IV)

11 mg of tetrazole

0.5 ml anhydrous CH_3CN

After 8 min the slide was washed with CH_3CN then oxidized with $\text{I}_2/\text{H}_2\text{O}/\text{THF}/\text{lutidine}$ for 1 min, washed, and then dried. The slide was then soaked for 30 min in a PBS/0.05% Tween 20 buffer and the solution then shaken off. The slide was next treated with FITC-labeled streptavidin at 10 $\mu\text{g}/\text{ml}$ in the same buffer system for 30 min. After this time the streptavidin-buffer system was rinsed off with fresh PBS/0.05% Tween 20 buffer and then the slide was finally agitated in distilled water for about 1/2 h. After drying, the slide was examined by fluorescence microscopy (see FIG. 2 and FIG. 3).

45

4. Substrate Preparation

Before attachment of reactive groups it is preferred to clean the substrate which is, in a preferred embodiment, a glass substrate such as a microscope slide or cover slip. A roughened surface will be useable but a plastic or other solid substrate is also appropriate. According to one embodiment the slide is soaked in an alkaline bath consisting of, e.g., 1 liter of 95% ethanol with 120 ml of water and 120 grams of sodium hydroxide for 12 hours. The slides are washed with a buffer and under running water, allowed to air dry, and rinsed with a solution of 95% ethanol.

The slides are then aminated with, e.g., aminopropyltriethoxysilane for the purpose of attaching amino groups to the glass surface on linker molecules, although other omega functionalized silanes could also be used for this purpose. In one embodiment 0.1% aminopropyltriethoxysilane is utilized, although solutions with concentrations from 10⁻⁷% to 10% may be used, with about 10⁻³% to 2% preferred. A

US 6,403,320 B1

105

0.1% mixture is prepared by adding to 100 ml of a 95% ethanol/5% water mixture, 100 microliters (μ l) of amino-propyltriethoxysilane. The mixture is agitated at about ambient temperature on a rotary shaker for an appropriate amount of time, e.g., about 5 minutes. 500 μ l of this mixture is then applied to the surface of one side of each cleaned slide. After 4 minutes or more, the slides are decanted of this solution and thoroughly rinsed three times or more by dipping in 100% ethanol.

After the slides dry, they are heated in a 110–120° C. vacuum oven for about 20 minutes, and then allowed to cure at room temperature for about 12 hours in an argon environment. The slides are then dipped into DMF (dimethylformamide) solution, followed by a thorough washing with methylene chloride.

5. Linker Attachment, Blocking of Free Sites

The aminated surface of the slide is then exposed to about 500 μ l of, for example, a 30 millimolar (mM) solution of NVOC-nucleotide-NHS (N-hydroxysuccinimide) in DMF for attachment of a NVOC-nucleotide to each of the amino groups. See, e.g., SIGMA Chemical Company for various nucleotide derivatives. The surface is washed with, for example, DMF, methylene chloride, and ethanol.

Any unreacted aminopropyl silane on the surface, i.e., those amino groups which have not had the NVOC-nucleotide attached, are now capped with acetyl groups (to prevent further reaction) by exposure to a 1:3 mixture of acetic anhydride in pyridine for 1 hour. Other materials which may perform this residual capping function include trifluoroacetic anhydride, formicacetic anhydride, or other reactive acylating agents. Finally, the slides are washed again with DMF, methylene chloride, and ethanol.

6. Synthesis of Eight Trimers of C and T

FIG. 4 illustrates a possible synthesis of the eight trimers of the two-monomer set: cytosine and thymine (represented by C and T, respectively). A glass slide bearing silane groups terminating in 6-nitroveratryloxycarboxamide (NVOC-NH) residues is prepared as a substrate. Active esters (pentafluorophenyl, OBt, etc.) of cytosine and thymine protected at the 5' hydroxyl group with NVOC are prepared as reagents. While not pertinent to this example, if side chain protecting groups are required for the monomer set, these must not be photoreactive at the wavelength of light used to protect the primary chain.

For a monomer set of size n, nxl cycles are required to synthesize all possible sequences of length l. A cycle consists of:

1. Irradiation through an appropriate mask to expose the 5'-OH groups at the sites where the next residue is to be added, with appropriate washes to remove the by-products of the deprotection.

2. Addition of a single activated and protected (with the same photochemically-removable group) monomer, which will react only at the sites addressed in step 1, with appropriate washes to remove the excess reagent from the surface.

The above cycle is repeated for each member of the monomer set until each location on the surface has been extended by one residue in one embodiment. In other embodiments, several residues are sequentially added at one location before moving on to the next location. Cycle times will generally be limited by the coupling reaction rate, now as short as about 10 min in automated oligonucleotide synthesizers. This step is optionally followed by addition of

Of course, greater diversity is obtained by using masking strategies which will also include the synthesis of polymers having a length of less than l. If, in the extreme case, all polymers having a length less than or equal to l are synthesized, the number of polymers synthesized will be:

106

$$n^l + n^{l-1} + \dots + n^1.$$

(3)

The maximum number of lithographic steps needed will generally be n for each "layer" of monomers, i.e., the total number of masks (and, therefore, the number of lithographic steps) needed will be nxl. The size of the transparent mask regions will vary in accordance with the area of the substrate available for synthesis and the number of sequences to be formed. In general, the size of the synthesis areas will be:

$$\text{size of synthesis areas} = (A)/(S)$$

where:

A is the total area available for synthesis; and

S is the number of sequences desired in the area.

It will be appreciated by those of skill in the art that the above method could readily be used to simultaneously produce thousands or millions of oligomers on a substrate using the photolithographic techniques disclosed herein. Consequently, the method results in the ability to practically test large numbers of, for example, di, tri, tetra, penta, hexa, hepta, octa, nona, deca, even dodecanucleotides, or larger polynucleotides (or correspondingly, polypeptides).

The above example has illustrated the method by way of a manual example. It will of course be appreciated that automated or semi-automated methods could be used. The substrate would be mounted in a flow cell for automated addition and removal of reagents, to minimize the volume of reagents needed, and to more carefully control reaction conditions. Successive masks will be applicable manually or automatically. See, e.g., U.S. Ser. No. 07/492,462 (VLSIPS CIP) and U.S. Ser. No. 07/624,120, (automated VLSIPS).

7. Labeling of Target

The target oligonucleotide can be labeled using standard procedures referred to above. As discussed, for certain situations, a reagent which recognizes interaction, e.g., ethidium bromide, may be provided in the detection step. Alternatively, fluorescence labeling techniques may be applied, see, e.g., Smith, et al. (1986) *Nature*, 321: 674–679; and Prober, et al. (1987) *Science*, 238:336–341. The techniques described therein will be followed with minimal modifications as appropriate for the label selected.

8. Dimers of A, C, G, and T

The described technique may be applied, with photosensitive blocked nucleotides corresponding to adenine, cytosine, guanine, and thymine, to make combinations of polynucleotides consisting of each of the four different nucleotides. All 16 possible dimers would be made using a minor modification of the described method.

9. 10-mers of A, C, G, and T

The described technique for making dimers of A, C, G, and T may be further extended to make longer oligonucleotides. The automated system described, e.g., in U.S. Ser. No. 07/492,462 (VLSIPS CIP), and U.S. Ser. No. 07/624,120, (automated VLSIPS), can be adapted to make all possible 10-mers composed of the 4 nucleotides A, C, G, and T. The photosensitive, blocked nucleotide analogues have been described above, and would be readily adaptable to longer oligonucleotides.

10. Specific Recognition Hybridization to 10-mers

The described hybridization conditions are directly applicable to the sequence specific recognition reagents attached to the substrate, produced as described immediately above. The 10-mers have an inherent property of hybridizing to a complementary sequence. For optimum discrimination between full matching and some mismatch, the conditions of hybridization should be carefully selected, as described above. Careful control of the conditions, and titration of

US 6,403,320 B1

107

parameters should be performed to determine the optimum collective conditions.

11. Hybridization

Hybridization conditions are described in detail, e.g., in Hames and Higgins (1985) *Nucleic Acid Hybridisation: A Practical Approach*; and the considerations for selecting particular conditions are described, e.g., in Wetmur and Davidson, (1988) *J. Mol. Biol.* 31:349–370, and Wood et al. (1985) *Proc. Natl. Acad. Sci. USA* 82:1585–1588. As described above, conditions are desired which can distinguish matching along the entire length of the probe from where there is one or more mismatched bases. The length of incubation and conditions will be similar, in many respects, to the hybridization conditions used in Southern blot transfers. Typically, the GC bias may be minimized by the introduction of appropriate concentrations of the alkylammonium buffers, as described above.

Titration of the temperature and other parameters is desired to determine the optimum conditions for specificity and distinguishability of absolutely matched hybridization from mismatched hybridization.

A fluorescently labeled target or set of targets are generated, as described in Prober, et al. (1987) *Science* 238:336–341, or Smith, et al. (1986) *Nature* 321:674–679. Preferably, the target or targets are of the same length as, or slightly longer, than the oligonucleotide probes attached to the substrate and they will have known sequences. Thus, only a few of the probes hybridize perfectly with the target, and which particular ones did would be known.

The substrate and probes are incubated under appropriate conditions for a sufficient period of time to allow hybridization to completion. The time is measured to determine when the probe-target hybridizations have reached completion. A salt buffer which minimizes GC bias is preferred, incorporating, e.g., buffer, such as tetramethyl ammonium or triethyl ammonium ion at between about 2.4 and 3.0 M. See Wood, et al. (1965) *Proc. Nat'l Acad. Sci. USA* 82:1585–1588. This time is typically at least about 30 min, and may be as long as about 1–5 days. Typically very long matches will hybridize more quickly, very short matches will hybridize less quickly, depending upon relative target and probe concentrations. The hybridization will be performed under conditions where the reagents are stable for that time duration.

Upon maximal hybridization, the conditions for washing are titrated. Three parameters initially titrated are time, temperature, and cation concentration of the wash step. The matrix is scanned at various times to determine the conditions at which the distinguishability between true perfect hybrid and mismatched hybrid is optimized. These conditions will be preferred in the sequencing embodiments.

12. Positional Detection of Specific Interaction

As indicated above, the detection of specific interactions may be performed by detecting the positions where the labeled target sequences are attached. Where the label is a fluorescent label, the apparatus described, e.g., in U.S. Ser. No. 07/492,462 (VLSIPS CIP); and U.S. Ser. No. 07/624,120, may be advantageously applied. In particular, the synthetic processes described above will result in a matrix pattern of specific sequences attached to the substrate, and a known pattern of interactions can be converted to corresponding sequences.

In an alternative embodiment, a separate reagent which differentially interacts with the probe and interacted probe/targets can indicate where interaction occurs or does not occur. A single-strand specific reagent will indicate where no interaction has taken place, while a double-strand specific

108

reagent will indicate where interaction has taken place. An intercalating dye, e.g., ethidium bromide, may be used to indicate the positions of specific interaction.

13. Analysis

Conversion of the positional data into sequence specificity will provide the set of subsequences whose analysis by overlap segments, may be performed, as described above. Analysis is provided by the methodology described above, or using, e.g., software available from the Genetic Engineering Center, P.O. Box 794, 11000 Belgrade, Yugoslavia (Yugoslav group). See, also, Macevitz, PCT publication no. WO 90/04652, which is hereby incorporated herein by reference.

B. Polypeptide

The description of the preparation of short peptides on a substrate incorporates by reference sections in U.S. Ser. No. 07/492,462 (VLSIPS CIP), and described below.

1. Slide Preparation

Preparation of the substrate follows that described above for nucleotides.

2. Linker Attachment, Blocking of Free Sites

The aminated surface of the slide is exposed to about 500 μ l of, e.g., a 30 millimolar (mM) solution of NVOC-GABA (gamma amino butyric acid) NHS (N-hydroxysuccinimide) in DMF for attachment of a NVOC-GABA to each of the amino groups. The surface is washed with, for example, DMF, methylene chloride, and ethanol. See U.S. Ser. No. 07/624,120, for details on amino acid chemistry.

Any unreacted aminopropyl silane on the surface, i.e., those amino groups which have not had the NVOC-GABA attached, are now capped with acetyl groups (to prevent further reaction) by exposure to a 1:3 mixture of acetic anhydride in pyridine for 1 hour. Other materials which may perform this residual capping function include trifluoroacetic anhydride, formicacetic anhydride, or other reactive acylating agents. Finally, the slides are washed again with DMF, methylene chloride, and ethanol.

3. Synthesis of 8 Trimers of "A" and "B"

See U.S. Ser. No. 07/492,462 (VLSIPS CIP) which describes the preparation of glycine and phenylalanine trimers. The technique is similar to the method described above for making trimers of C and T, but substituting photosensitive blocked glycine for the C derivative and photosensitive blocked phenylalanine for the T derivative.

4. Synthesis of a Dimer of an Aminopropyl Group and a Fluorescent Group

In synthesizing the dimer of an aminopropyl group and a fluorescent group, a functionalized durapore membrane was used as a substrate. The Durapore membrane was a polyvinylidene difluoride with aminopropyl groups. The aminopropyl groups were protected with the DDZ group by reaction of the carbonyl chloride with the amino groups, a reaction readily known to those of skill in the art. The surface bearing these groups was placed in a solution of THF and contacted with a mask bearing a checkerboard pattern of 1 mm opaque and transparent regions. The mask was exposed to ultraviolet light having a wavelength down to at least about 280 nm for about 5 minutes at ambient temperature, although a wide range of exposure times and temperatures may be appropriate in various embodiments of the invention. For example, in one embodiment, an exposure time of between about 1 and 5000 seconds may be used at process temperatures of between -70 and $+50^{\circ}$ C.

In one preferred embodiment, exposure times of between about 1 and 500 seconds at about ambient pressure are used. In some preferred embodiments, pressure above ambient is used to prevent evaporation.

US 6,403,320 B1

109

The surface of the membrane was then washed for about 1 hour with a fluorescent label which included an active ester bound to a chelate of a lanthanide. Wash times will vary over a wide range of values from about a few minutes to a few hours. These materials fluoresce in the red and the green visible region. After the reaction with the active ester in the fluorophore was complete, the locations in which the fluorophore was bound could be visualized by exposing them to ultraviolet light and observing the red and the green fluorescence. It was observed that the derivatized regions of the substrate closely corresponded to the original pattern of the mask.

5. Demonstration of Signal Capability

Signal detection capability was demonstrated using a low-level standard fluorescent bead kit manufactured by Flow Cytometry Standards and having model no. 824. This kit includes 5.8 μm diameter beads, each impregnated with a known number of fluorescein molecules.

One of the beads was placed in the illumination field on the scan stage in a field of a laser spot which was initially shuttered. After being positioned in the illumination field, the photon detection equipment was turned on. The laser beam was unblocked and it interacted with the particle bead, which then fluoresced. Fluorescence curves of beads impregnated with 7,000 and 29,000 fluorescein molecules, are shown in FIGS. 11A and 11B, respectively of U.S. Ser. No. 07/492,462 (VLSIPS CIP). On each curve, traces for beads without fluorescein molecules are also shown. These experiments were performed with 488 nm excitation, with 100 μW of laser power. The light was focused through a 40 power 0.75 NA objective.

The fluorescence intensity in all cases started off at a high value and then decreased exponentially. The fall-off in intensity is due to photobleaching of the fluorescein molecules. The traces of beads without fluorescein molecules are used for background subtraction. The difference in the initial exponential decay between labeled and nonlabeled beads is integrated to give the total number of photon counts, and this number is related to the number of molecules per bead. Therefore, it is possible to deduce the number of photons per fluorescein molecule that can be detected. This calculation indicates the radiation of about 40 to 50 photons per fluorescein molecule are detected.

6. Determination of the Number of Molecules Per Unit Area

Aminopropylated glass microscope slides prepared according to the methods discussed above were utilized in order to establish the density of labeling of the slides. The free amino termini of the slides were reacted with FITC (fluorescein isothiocyanate) which forms a covalent linkage with the amino group. The slide is then scanned to count the number of fluorescent photons generated in a region which, using the estimated 40–50 photons per fluorescent molecule, enables the calculation of the number of molecules which are on the surface per unit area.

A slide with aminopropyl silane on its surface was immersed in a 1 mM solution of FITC in DMF for 1 hour at about ambient temperature. After reaction, the slide was washed twice with DMF and then washed with ethanol, water, and then ethanol again. It was then dried and stored in the dark until it was ready to be examined.

Through the use of curves similar to those shown in FIG. 11 of U.S. Ser. No. 07/492,462 (VLSIPS CIP), and by integrating the fluorescent counts under the exponentially decaying signal, the number of free amino groups on the surface after derivitization was determined. It was determined that slides with labeling densities of 1 fluorescein per $10^3 \times 10^3$ to $\sim 2 \times 2$ nm could be reproducibly made as the

110

concentration of aminopropyltriethoxysilane varied from $10^{-5}\%$ to $10^{-1}\%$.

7. Removal of NOVC and Attachment of a Fluorescent Marker

5 NVOC-GABA groups were attached as described above. The entire surface of one slide was exposed to light so as to expose a free amino group at the end of the gamma amino butyric acid. This slide, and a duplicate which was not exposed, were then exposed to fluorescein isothiocyanate (FITC).

10 FIG. 12A of U.S. Ser. No. 07/492,462 (VLSIPS CIP) illustrates the slide which was not exposed to light, but which was exposed to FITC. The units of the x axis are time and the units of the y axis are counts. The trace contains a certain amount of background fluorescence. The duplicate slide was exposed to 350 nm broadband illumination for about 1 minute (12 mW/cm², ~ 350 nm illumination), washed and reacted with FITC. A large increase in the level of fluorescence is observed, which indicates photolysis has exposed a number of amino groups on the surface of the slides for attachment of a fluorescent marker.

8. Use of a Mask in Removal of NVOC

The next experiment was performed with a 0.1% amino-propylated slide. Light from a Hg—Xe arc lamp was imaged onto the substrate through a laser-ablated chrome-on-glass mask in direct contact with the substrate.

This slide was illuminated for approximately 5 minutes, with 12 mW of 350 nm broadband light and then reacted with the 1 mM FITC solution. It was put on the laser detection scanning stage and a graph was plotted as a two-dimensional representation of position color-coded for fluorescence intensity. The experiment was repeated a number of times through various masks. The fluorescence patterns for a $100 \times 100 \mu\text{m}$ mask, a $50 \mu\text{m}$ mask, a $20 \mu\text{m}$ mask, and a $10 \mu\text{m}$ mask indicate that the mask pattern is distinct down to at least about $10 \mu\text{m}$ squares using this lithographic technique.

9. Attachment of YGGFL and Subsequent Exposure to Herz Antibody and Goat Anti-mouse Antibody

40 In order to establish that receptors to a particular polypeptide sequence would bind to a surface-bound peptide and be detected, Leu enkephalin was coupled to the surface and recognized by an antibody. A slide was derivatized with 0.1% amino propyl-triethoxysilane and protected with NVOC. A $500 \mu\text{m}$ checkerboard mask was used to expose the slide in a flow cell using backside contact printing. The Leu enkephalin sequence (H_2N -tyrosine, glycine, glycine, phenylalanine, leucine-COOH, otherwise referred to herein as YGGFL) was attached via its carboxy end to the exposed amino groups on the surface of the slide. The peptide was added in DMF solution with the BOP/HOBT/DIEA coupling reagents and recirculated through the flow cell for 2 hours at room temperature.

A first antibody, known as the Herz antibody, was applied to the surface of the slide for 45 minutes at $2 \mu\text{g}/\text{ml}$ in a supercocktail (containing 1% BSA and 1% ovalbumin also in this case). A second antibody, goat anti-mouse fluorescein conjugate, was then added at $2 \mu\text{g}/\text{ml}$ in the supercocktail buffer, and allowed to incubate for 2 hours.

60 The results of this experiment were plotted as fluorescence intensity as a function of position. This image was taken at $10 \mu\text{m}$ steps and showed that not only can deprotection be carried out in a well defined pattern, but also that (1) the method provided for successful coupling of peptides to the surface of the substrate, (2) the surface of a bound peptide was available for binding with an antibody, and (3) that the detection apparatus-capabilities were sufficient to

detect binding of a receptor. Moreover, the Herz antibody is a sequence specific reagent which may be used advantageously as a sequence specific recognition reagent. It may be used, if specificity is high, for sequencing purposes, and, at least, for fingerprinting and mapping uses.

10. Monomer-by-monomer Formation of YGGFL and Subsequent Exposure to Labeled Antibody

Monomer-by-monomer synthesis of YGGFL and GGFL in alternate squares was performed on a slide in a checkerboard pattern and the resulting slide was exposed to the Herz antibody.

A slide is derivatized with the aminopropyl group, protected in this case with t-BOC (t-butoxycarbonyl). The slide was treated with TFA to remove the t-BOC protecting group. E-aminocaproic acid, which was t-BOC protected at its amino group, was then coupled onto the aminopropyl groups. The aminocaproic acid serves as a spacer between the aminopropyl group and the peptide to be synthesized. The amino end of the spacer was deprotected and coupled to UVOC-leucine. The entire slide was then illuminated with 12 mW of 325 nm broadband illumination. The slide was then coupled with NVOC-phenylalanine and washed. The entire slide was again illuminated, then coupled to NVOC-glycine and washed. The slide was again illuminated and coupled to NVOC-glycine to form the sequence shown in the last portion of FIG. 13A of U.S. Ser. No. 07/492,462 (VLSIPS CIP).

Alternating regions of the slide were then illuminated using a projection print using a 500x500 μ m checkerboard mask; thus, the amino group of glycine was exposed only in the lighted areas. When the next coupling chemistry step was carried out, NVOC-tyrosine was added, and it coupled only at those spots which had received illumination. The entire slide was then illuminated to remove all the NVOC groups, leaving a checkerboard of YGGFL in the lighted areas and in the other areas, GGFL. The Herz antibody (which recognizes the YGGFL, but not GGFL) was then added, followed by goat anti-mouse fluorescein conjugate.

The resulting fluorescence scan showed dark areas containing the tetrapeptide GGFL, which is not recognized by the Herz antibody (and thus there is no binding of the goat anti-mouse antibody with fluorescein conjugate), and red areas in which YGGFL was present. The YGGFL pentapeptide is recognized by the Herz antibody and, therefore, there is antibody in the lighted regions for the fluorescein-conjugated goat anti-mouse to recognize.

Similar patterns for a 50 μ m mask used in direct contact ("proximity print") with the substrate provided a pattern which was more distinct and the corners of the checkerboard pattern were touching as a result of the mask being placed in direct contact with the substrate (which reflects the increase in resolution using this technique).

11. Monomer-by-monomer Synthesis of YGGFL and PGGFL

A synthesis using a 50 μ m checkerboard mask was conducted. However, P was added to the GGFL sites on the substrate through an additional coupling step. P was added by exposing protected GGFL to light through a mask, and subsequence exposure to P in the manner set forth above. Therefore, half of the regions on the substrate contained YGGFL and the remaining half contained PGGFL.

The fluorescence plot for this experiment showed the regions are again readily discernable between those in which binding did and did not occur. This experiment demonstrated that antibodies are able to recognize a specific sequence and that the recognition is not length-dependent. ps 12. Monomer-by-monomer Synthesis of YGGFL and YPGGFL

In order to further demonstrate the operability of the invention, a 50 μ m checkerboard pattern of alternating YGGFL and YPGGFL was synthesized on a substrate using techniques like those set forth above. The resulting fluorescence plot showed that the antibody was clearly able to recognize the YGGFL sequence and did not bind significantly at the YPGGFL regions.

13. Synthesis of an Array of Sixteen Different Amino Acid Sequences and Estimation of Relative Binding Affinity to Herz Antibody

Using techniques similar to those set forth above, an array of 16 different amino acid sequences (replicated four times) was synthesized on each of two glass substrates. The sequences were synthesized by attaching the sequence NVOC-GFL across the entire surface of the slides. Using a series of masks, two layers of amino acids were then selectively applied to the substrate. Each region had dimensions of 0.25 cmx0.0625 cm. The first slide contained amino acid sequences containing only L-amino acids while the second slide contained selected D-amino acids. Various regions on the first and second slides, were duplicated four times on each slide. The slides were then exposed to the Herz antibody and fluorescein-labeled goat anti-mouse antibodies.

A fluorescence plot of the first slide, which contained only L-amino acids showed red areas (indicating strong binding, i.e., 149,000 counts or more) and black areas (indicating little or no binding of the Herz antibody, i.e., 20,000 counts or less). The sequence YGGFL was clearly most strongly recognized. The sequences YAGFL and YSGFL also exhibited strong recognition of the antibody. By contrast, most of the remaining sequences showed little or no binding. The four duplicate portions of the slide were extremely consistent in the amount of binding shown therein.

A fluorescence plot of the D-amino acid slide indicated that strongest binding was exhibited by the YGGFL sequence. Significant binding was also detected to YaGFL, YsGFL, and YpGFL. The remaining sequences showed less binding with the antibody. Low binding efficiency of the sequence yGGFL was observed.

Table 6 lists the various sequences tested in order of relative fluorescence, which provides information regarding relative binding affinity.

TABLE 6

Apparent Binding to Herz Ab	
L- a.a. Set	D- a.a. Set
YGGFL	YGGFL
YAGFL	YaGFL
YSGFL	YsGFL
LGGFL	YpGFL
FGGFL	fGGFL
YPGFL	yGGFL
LAGFL	faGFL
FAGFL	wGGFL
WGGFL	yaGFL
	fpGFL
	waGFL

14. Illustrative Alternative Embodiment

According to an alternative embodiment of the invention, the methods provide for attaching to the surface a caged binding member which, in its caged form, has a relatively low affinity for other potentially binding species, such as receptors and specific binding substances. Such techniques are more fully described in copending application Ser. No. 25 404,920, filed Sep. 8, 1989, and incorporated herein by

US 6,403,320 B1

113

reference for all purposes. See also U.S. Ser. No. 07/435,316 (caged biotin parent) and U.S. Ser. No. 07/612,671 (caged biotin CIP), each of which is hereby incorporated herein by reference.

According to this alternative embodiment, the invention provides methods for forming predefined regions on a surface of a solid support, wherein the predefined regions are capable of immobilizing receptors. The methods make use of caged binding members attached to the surface to enable selective activation of the predefined regions. The caged binding members are liberated to act as binding members ultimately capable of binding receptors upon selective activation of the predefined regions. The activated binding members are then used to immobilize specific molecules such as receptors on the predefined region of the surface. The above procedure is repeated at the same or different sites on the surface so as to provide a surface prepared with a plurality of regions on the surface containing, for example, the same or different receptors. When receptors immobilized in this way have a differential affinity for one or more ligands, screenings and assays for the ligands can be conducted in the regions of the surface containing the receptors.

The alternative embodiment may make use of novel caged binding members attached to the substrate. Caged (unactivated) members have a relatively low affinity for receptors of substances that specifically bind to uncaged binding members when compared with the corresponding affinities of activated binding members. Thus, the binding members are protected from reaction until a suitable source of energy is applied to the regions of the surface desired to be activated. Upon application of a suitable energy source, the caging groups labilize, thereby presenting the activated binding member. A typical energy source will be light.

Once the binding members on the surface are activated they may be attached to a receptor. The receptor chosen may be a monoclonal antibody, a nucleic acid sequence, a drug receptor, etc. The receptor will usually, though not always, be prepared so as to permit attaching it, directly or indirectly, to a binding member. For example, a specific binding substance having a strong binding affinity for the binding member and a strong affinity for the receptor or a conjugate of the receptor may be used to act as a bridge between binding members and receptors if desired. The method uses a receptor prepared such that the receptor retains its activity toward a particular ligand.

Preferably, the caged binding member attached to the solid substrate will be a photoactivatable biotin complex, i.e., a biotin molecule that has been chemically modified with photoactivatable protecting groups so that it has a significantly reduced binding affinity for avidin or avidin analogs than does natural biotin. In a preferred embodiment, the protecting groups localized in a predefined region of the surface will be removed upon application of a suitable source of radiation to give binding members, that are biotin or a functionally analogous compound having substantially the same binding affinity for avidin or avidin analogs as does biotin.

In another preferred embodiment, avidin or an avidin analog is incubated with activated binding members on the surface until the avidin binds strongly to the binding members. The avidin so immobilized on predefined regions of the surface can then be incubated with a desired receptor or conjugate of a desired receptor. The receptor will preferably be biotinylated, e.g., a biotinylated antibody, when avidin is immobilized on the predefined regions of the surface. Alternatively, a preferred embodiment will present an avidin/biotinylated receptor complex, which has been previously prepared, to activated binding members on the surface.

114

II. Fingerprinting

The above section on generation of reagents for sequencing provides specific reagents useful for fingerprinting applications. Fingerprinting embodiments may be applied towards polynucleotide fingerprinting, polypeptide fingerprinting, cell and tissue classification, cell and tissue temporal development stage classification, diagnostic tests, forensic uses for individual identification, classification of organisms, and genetic screening of individuals. Mapping applications are also described below.

A. Polynucleotide Fingerprint

Polynucleotide fingerprinting may use reagents similar to those described above for probing a sequence for the presence of specific subsequences found therein. Typically, the subsequences used for fingerprinting will be longer than the sequences used in oligonucleotide sequencing. In particular, specific long segments may be used to determine the similarity of different samples of nucleic acids. They may also be used to fingerprint whether specific combinations of information are provided therein. Particular probe sequences are selected and attached in a positional manner to a substrate. The means for attachment may be either using a caged biotin method described, e.g., in U.S. Ser. No. 07/612,671 (caged biotin CIP), or by another method using targeting molecules. For example, a short polypeptide of specific sequence may be attached to an oligonucleotide and targeted to specific positions on a substrate having antibodies attached thereto, the antibodies exhibiting specificity for binding to those short peptide sequences. In another embodiment, an unnatural nucleotide or similar complementary binding molecule may be attached to the fingerprinting probe and the probe thereby directed towards complementary sequences on a VLSIPS substrate. Typically, unnatural nucleotides would be preferred, e.g., unnatural optical isomers, which would not interfere with natural nucleotide interactions.

Having produced a substrate with particular fingerprint probes attached thereto at positionally defined regions, the substrate may be used in a manner quite similar to the sequencing embodiment to provide information as to whether the fingerprint probes are detecting the corresponding sequence in a target sequence. This will often provide information similar to a Southern blot hybridization.

B. Polypeptide Fingerprint

A polypeptide fingerprint may be performed using antibodies which recognize specific antigens on the polypeptide. For example, monoclonal antibodies which recognize specific sequences or antigens on a polypeptide may be used to determine whether those epitopes are found on a particular protein. For example, particular patterns of epitopes would be found on various types of proteins. This will lead to the discovery that specific epitopes, or antigenic determinants, which are characteristic of, e.g., beta sheet segments, will be identified as will particular different types of domains in various protein types. Thus, a screening method may be devised which can classify polypeptides, either native or denatured, into various new classes defined by the epitopes existing thereon.

In addition, once the substrate is generated in the manners described above, a target peptide is exposed to the substrate. The target may be either native or denatured, though the conditions used to denature the polypeptide may interfere with the specific interaction between the polypeptide and the recognition reagent. This method is not dependent on the fact that the polypeptide is a single chain, thus protein complexes may also be fingerprinted using this methodology. Structures such as multi-subunit proteins, associations of proteins, ribosomes, nucleosomes, and other small cellu-

US 6,403,320 B1

115

lar structures may also be fingerprinted and classified according to the presence of specific recognizable features thereon.

Peptide fingerprinting may be useful, for example, in correlating with particular physiological conditions or developmental stages of a cell or organism. Thus, a biological sample may be fingerprinted to determine the presence in that sample of a plurality of different polypeptides which are each individually fingerprinted. In an alternative embodiment, a polypeptide itself is not fingerprinted but a biological sample is fingerprinted searching for specific epitopes, e.g., polypeptide, carbohydrate, nucleic acid, or any of a number of other specific recognizable structural features.

The conditions for the interactions using antibodies is described, e.g., in Harlow and Lane (1988) *Antibodies: A Laboratory Manual*, Cold Spring Harbor Press, New York. The conditions should be titrated for temperature, buffer composition, time, and other important parameters in an antibody interaction.

C. Cell Classification Scheme

The present invention can be used for cell classification using fingerprinting type technology as described above in the polypeptide fingerprint. Classes of cells are typically defined by the presence of common functions which are usually reflected by structural features. Thus, a plant cell is classified differently from an animal cell by a number of structural features. Given an unknown cell, the present invention provides improved means for distinguishing the different cell types. Once a cell classification scheme is developed and the structural features which define it are identified using the present invention, homogeneous cell population expressing these features may be separated from others. Standard cell sorters may be coupled with recognition reagents and labels which can distinguish various cell types.

a. T-Cell Classes

T-cell classes are defined on the basis of expression of particular antigens characteristic of each class. For example, mouse T-cell differentiation markers include the LY antigens. With the plurality of different antigens which may be tested using antibody or other recognition reagents, new populations and classes of cells may be defined. For example, different neural cell types may be defined on the basis of cell surface antigens. Different tissue types will be defined on the basis of tissue specific antigens. Developmental cell classes will be similarly defined. All of these screenings can make use of the VLSIPS substrates with specific recognition molecules attached thereto. The substrates are exposed to the cell types directly, assaying for attachment of cells to specific regions, or are exposed to products of a population of cells, e.g., a supernatant, or a cell lysate.

Once a cell classification scheme has been correlated with specific structural markers therein, reagents which recognize those features may be developed and used in a fluorescence activated cell sorter as described, e.g., in Dangel, J. and Herzenberg (1982) *J. Immunological Methods* 52:

1-14; and Becton Dickinson, Fluorescence Activated Cell Sorters Division, San Jose, Calif. This will provide a homogeneous population of cells whose function has been defined by structure.

b. B-Cell Classes

The present cell classification scheme may also be used to determine specific B-cell classes. For example, B-cells specific for producing IgM, IgG, IgD, IgE, and IgA may be defined by the internal expression of specific mRNA

116

sequences encoding each type of immunoglobulin. The classification scheme may depend on either extracellularly expressed markers which are correlated as being diagnostic of specific stages in development, or intracellular mRNA sequences which indicate particular functions.

D. Temporal Development Scheme

1. Developmental Antigens

The present fingerprinting invention also allows cell classification by expression of developmental antigens. For example, a lymphocyte stem cell expresses a particular combination of antigens. As the lymphocyte develops through a program developmental scheme, at various stages it expresses particular antigens which are diagnostic of particular stages in development. Again, the fingerprinting methodology allows for the definition of specific structural features which are diagnostic of developmental or functional features which will allow classification of cells into temporal developmental classes. Cells, products of those cells, or lysates of those cells will be assayed to determine the developmental stage of the source cells. In this manner, once a developmental stage is defined, specific synchronized populations of cells will be selected out of another population. These synchronized populations may be very important in determining the biological mechanisms of development.

2. Developmental mRNA Expression

Besides expressed antigens, the present invention also allows for fingerprinting of the mRNA population of a cell. In this fashion, the mRNA population, which should be a good determinant of developmental stage, will be correlated with other structural features of the cell. In this manner, cells at specific developmental stages will be characterized by the intracellular environment, as well as the extracellular environment. The present invention also allows the combination of definitions based, in part, upon antigens and, in part, upon mRNA expression.

In one embodiment, the two may be combined in a single incubation step. A particular incubation condition may be found which is compatible with both hybridization recognition non-hybridization recognition molecules. Thus, e.g., an incubation condition may be selected which allows both specificity of antibody binding and specificity of nucleic acid hybridization. This allows simultaneous performance of both types of interactions on a single matrix. Again, where developmental mRNA patterns are correlated with structural features, or with probes which are able to hybridize to intracellular mRNA populations, a cell sorter may be used to sort specifically those cells having desired mRNA population patterns.

E. Diagnostic Tests

The present invention also provides the ability to perform diagnostic tests. Diagnostic tests typically are based upon a fingerprint type assay, which tests for the presence of specific diagnostic structural features. Thus, the present invention provides means for viral strain identification, bacterial strain identification, and other diagnostic tests using positionally defined specific reagents. The present invention also allows for determining a spectrum of allergies, diagnosing a biological sample for any or all of the above, and testing for many other conditions.

1. Viral Identification

The present invention provides reagents and methodology for identifying viral strains. The specific reagents may be either antibodies or recognition proteins which bind to specific viral epitopes preferably surface exposed, but may make use of internal epitopes, e.g., in a denatured viral sample. In an alternative embodiment, the viral genome may be probed for specific sequences which are characteristic of

US 6,403,320 B1

117

particular viral strains. As above, a combination of the two may be performed simultaneously in a single interaction step, or in separate tests, e.g., for both genetic characteristics and epitope characteristics.

2. Bacterial Identification

Similar techniques will be applicable to identifying a bacterial source. This may be useful in diagnosing bacterial infections, or in classifying sources of particular bacterial species. For example, the bacterial assay may be useful in determining the natural range of survivability of particular strains of bacteria across regions of the country or in different ecological niches.

3. Other Microbiological Identifications

The present invention provides means for diagnosis of other microbiological and other species, e.g., protozoal species and parasitic species in a biological sample, but also provides the means for assaying a combination of different infections. For example, a biological specimen may be assayed for the presence of any or all of these microbiological species. In human diagnostic uses, typical samples will be blood, sputum, stool, urine, or other samples.

4. Allergy Tests

An immobilized set of antigens may be attached to a solid substrate and, instead of the standard skin reaction tests, a blood sample may be assayed on such a substrate to determine the presence of antibodies, e.g., IgE or other type antibodies, which may be diagnostic of an allergic or immunological susceptibility. A standard radioallergosorbent test (RAST) may be used to check a much larger population of antigens.

In addition, an allergy like test may be used to diagnose the immunological history of a particular individual. For example, by testing the circulating antibodies in a blood sample, which reflects the immunological history and memory of an individual, it may be determined what infections may not have been historically presented to the immune system. In this manner, it may be possible to specifically supplement an immune system for a short period of time with IgG fractions made up of specific types of gamma globulins. Thus, hepatitis gamma globulin injections may be better designed for a particular environment which a person is expected to be exposed. This also provides the ability to identify genetically equivalent individuals who have immunologically different experiences. Thus, a blood sample from an individual who has a particular combination of circulating antibodies will likely be different from the combination of circulating antibodies found in a genetically similar or identical individual. This could allow for the distinction between clones of particular animals, e.g., mice, rats, or other animals.

F. Individual Identification

The present invention provides the ability to fingerprint and identify a genetic individual. This individual may be a bacterial or lower microorganism, as described above in diagnostic tests, or of a plant or animal. An individual may be identified genetically or immunologically, as described.

1. Genetic

Genetic fingerprinting has been utilized in comparing different related species in Southern hybridization blots. Genetic fingerprinting has also been used in forensic studies, see, e.g., Morris et al. (1989) *J. Forensic Science* 34: 1311-1317, and references cited therein. As described above, an individual may be identified genetically by a sufficiently large number of probes. The likelihood that another individual would have an identical pattern over a sufficiently large number of probes may be statistically negligible. However, it is often quite important that a large

118

number of probes be used where the statistical probability of matching is desired to be particularly low. In fact, the probes will optimally be selected for having high heterogeneity among the population. In addition, the fingerprint method may make use of the pattern of homologies indicated by a series of more and more stringent washes. Then, each position has both a sequence specificity and a homology measurement, the combination of which greatly increases the number of dimensions and the statistical likelihood of a perfect pattern match with another genetic individual.

2. Immunological

As indicated above in the diagnostic tests, it is possible to identify a particular immune system within a genetically homogeneous class of organisms by virtue of her immunological history. For example, a large colony of cloned mice may be distinguishable by virtue of each immunological history. For example, one mouse may have had an immunological response to exposure to antigen A to which her genetically identical sibling may have not been exposed. By virtue of this differential history, the first of the pair will likely have a high antibody titer against the antigen A whereas her genetically identical sibling will have not had a response to that antigen by virtue of never having been exposed to it. For this reason, immune systems may be identified by their immunological memories. Thus, immunological experience may also be a means for identifying a particular individual at a particular moment in her lifetime.

This same immunological screening may be used for other sorts of identifiable biological products. For example, an individual may be identified by her combination of expressed proteins. These proteins may reflect a physiological state of the individual, and would thus be useful in certain circumstances where diagnostic tests may be performed. For example, an individual may be identified, in part, by the presence of particular metabolic products.

In fact, a plant origin may be determined by virtue of having within its genome an unnatural sequence introduced to it by genetic breeders. Thus, a marker nucleic acid sequence may be introduced as a means to determine whether a genetic strain of a plant or animal originated from another particular source.

G. Genetic Screening

1. Test Alleles With Markers

The present invention provides for the ability to screen for genetic variations of individuals. For example, a number of genetic diseases are linked with specific alleles. See, e.g., Scriber, C. et al. (eds.) (1989) *The Metabolic Bases of Inherited Disease*, McGraw-Hill, New York. In one embodiment, cystic fibrosis has been correlated with a specific gene, see, Gregory et al. (1990) *Nature* 347: 382-386. A number of alleles are correlated with specific genetic deficiencies. See, e.g., McKusick, V. (1990) *Genetic Inheritance in Man: Catalogs of Autosomal Dominant, Autosomal Recessive, and X-linked Phenotypes*, Johns Hopkins University Press, Baltimore; Ott, J. (1985) *Analysis of Human Genetic Linkage*, Johns Hopkins University Press, Baltimore; Track, R. et al. (1989) *Banbury Report 32: DNA Technology and Forensic Science*, Cold Spring Harbor Press, New York; each of which is hereby incorporated herein by reference.

2. Amniocentesis

Typically, amniocentesis is used to determine whether chromosome translocations have occurred. The mapping procedure may provide the means for determining whether these translocations have occurred, and for detecting particular alleles of various markers.

US 6,403,320 B1

119

III. Mapping

A. Positionally Located Clones

The present invention allows for the positional location of specific clones useful for mapping. For example, caged biotin may be used for specifically positioning a probe to a location on a matrix pattern.

In addition, the specific probes may be positionally directed to specific locations on a substrate by targeting. For example, polypeptide specific recognition reagents may be attached to oligonucleotide sequences which can be complementarily targeted to specific locations on a VLSIPS substrate. Hybridization conditions, as applied for oligonucleotide probes, will be used to target the reagents to locations on a substrate having complementary oligonucleotides synthesized thereon. In another embodiment, oligonucleotide probes may be attached to specific polypeptide targeting reagents such as an antigen or antibody. These reagents can be directed towards a complementary antigen or antibody already attached to a VLSIPS substrate.

In another embodiment, an unnatural nucleotide which does not interfere with natural nucleotide complementary hybridization may be used to target oligonucleotides to particular positions on a substrate. Unnatural optical isomers of natural nucleotides should be ideal candidates.

In this way, short probes may be used to determine the mapping of long targets or long targets may be used to map the position of shorter probes. See, e.g., Craig et al. 1990 *Nuc. Acids Res.* 18: 2653–2660.

B. Positionally Defined Clones

Positionally defined clones may be transferred to a new substrate by either physical transfer or by synthetic means. Synthetic means may involve either a production of the probe on the substrate using the VLSIPS synthetic methods, or may involve the attachment of a targeting sequence made by VLSIPS synthetic methods which will target that positionally defined clone to a position on a new substrate. Both methods will provide a substrate having a number of positionally defined probes useful in mapping.

IX. Conclusion

The present inventions provide greatly improved methods and apparatus for synthesis of polymers on substrates. It is to be understood that the above description is intended to be illustrative and not restrictive. Many embodiments will be apparent to those of skill in the art upon reviewing the above description. By way of example, the invention has been described primarily with reference to the use of photoremovable protective groups, but it will be readily recognized by those of skill in the art that sources of radiation other than light could also be used. For example, in some embodiments it may be desirable to use protective groups which are sensitive to electron beam irradiation, x-ray irradiation, in combination with electron beam lithograph, or x-ray lithography techniques. Alternatively, the group could be removed by exposure to an electric current. The scope of the invention should, therefore, be determined not with reference to the above description, but should instead be determined with reference to the appended claims, along with the full scope of equivalents to which such claims are entitled.

All publications and patent applications referred to herein are incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually incorporated by reference. The present invention now being fully described, it will be apparent to one of ordinary skill in the art that many changes and modifications can be made thereto without departing from the spirit or scope of the appended claims.

120

What is claimed is:

1. A method of analyzing a receptor bound to an array of nucleic acids, comprising:

providing an array of different nucleic acids immobilized on a substrate, the different nucleic acids occupying different localized areas on the substrate of less than 0.01 cm^2 , at least some of the nucleic acids being bound to a fluorescently labeled receptor; and

detecting the localized areas of the array with a detector to detect light fluoresced from the labeled receptor bound to the nucleic acids in the localized areas, the detection step being one second or less, to detect which nucleic acids bind to the labeled receptor.

2. The method of claim 1, wherein the detector is a charge coupled device.

3. The method of claim 1, further comprising exposing the array to ultraviolet light.

4. The method of claim 1, wherein each location has an area of less than 10,000 square microns.

5. The method of claim 1, wherein each location has an area of less than 100 square microns.

6. The method of claim 1, wherein the array comprises at least 1000 probes.

7. The method of claim 1, wherein the combined area occupied by the at least 1000 nucleic acids is less than 1 cm^2 .

8. The method of claim 1, wherein the array comprises at least 10^6 nucleic acids per cm^2 .

9. The method of claim 1, wherein the nucleic acids have lengths of 2–100 nucleotides.

10. The method of claim 1, wherein the labeled receptor is a labeled nucleic acid.

11. The method of claim 1, wherein the labeled receptor is a labeled polypeptide.

12. The method of claim 1, wherein the known locations are squares or rectangles.

13. The method of claim 1, wherein the method further comprises:

directing light from a light source at the surface of the substrate; and storing a pattern of fluoresced light from the different localized areas in a data storage system.

14. The method of claim 1, wherein the method further comprises translating the detector relative to the substrate.

15. The method of claim 1, further comprising analyzing the stored pattern of fluoresced light to determine which nucleic acids bind to the receptor.

16. The method of claim 1, wherein the array comprises more than 10 different nucleic acids.

17. The method of claim 1, wherein the array comprises more than 100 different nucleic acids.

18. The method of claim 1, wherein the array comprises more than 10,000 different nucleic acids.

19. The method of claim 1, wherein the array comprises more than 100,000 different nucleic acids.

20. The method of claim 1, wherein the array comprises more than 1000 different nucleic acids at a density of more than 1000 nucleic acids per cm^2 .

21. The method of claim 1, wherein the array comprises more than 10,000 different nucleic acids at a density of more than 10,000 nucleic acids per cm^2 .

22. The method of claim 1, wherein the array comprises more than 1,000,000 different nucleic acids at a density of more than 1,000,000 nucleic acids per cm^2 .

23. The method of claim 1, further comprising:

illuminating the array with a device comprising:

a point light source to generate an excitation light;
an objective lens to focus the point light source at the surface of the substrate, whereby the fluorescently

US 6,403,320 B1

121

marked localized areas emit a fluoresced light in response to the excitation light;
 a dichroic mirror to reflect light having a wavelength of the excitation light and passing light having a wavelength of the fluoresced light;
 a photomultiplier and photon counter to detect the fluoresced light;
 moving the substrate relative to the excitation light with an x-y translation stage; and
 recording the fluoresced light as a function of a position on the surface of the substrate from which the fluoresced light was emitted with an appropriately programmed computer.

24. The method of claim 1, further comprising washing the surface of the substrate to remove unbound labeled receptor.

25. An apparatus for analyzing ligand-receptor binding, comprising:
 a substrate that comprises more than 1000 different ligands of known sequence collectively occupying less than 1 cm², the different ligands occupying different localized areas within the 1 cm²; and
 a detector capable of detecting light from fluorescently marked localized areas on the substrate.

26. The apparatus of claim 25, wherein the ligands are nucleic acids.

27. The apparatus of claim 25, wherein the ligands are polypeptides.

28. The apparatus of claim 25, wherein the ligands have known sequences.

29. An apparatus for analyzing ligand-receptor binding, comprising:
 a substrate that comprises at least 100 different ligands, collectively occupying less than 1 cm², the different ligands occupying different localized areas; and
 a detector to detect a label bound to said ligands at said localized areas.

30. The apparatus of claim 29, wherein the different ligands are nucleic acids.

31. The apparatus of claim 29, wherein the different ligands have known sequences.

32. The apparatus of claim 29, wherein the different ligands are peptides.

33. The apparatus of claim 29, wherein the marker is a fluorescent marker.

34. The apparatus of claim 29, wherein the marker is a radiolabelled marker.

35. An apparatus for detecting binding of nucleic acids; comprising:
 (a) a substrate having a surface comprising more than 10 different nucleic acids, having known sequences, at localized areas on the surface of the substrate, each of the localized areas having an area of 10⁻² cm² or less, at least one of the nucleic acids being bound to a labeled receptor;
 (b) an excitation light source;
 (c) a detector capable of receiving a signal from fluorescently marked localized areas on the substrate;
 (d) a translator to move the substrate relative to the detector; and
 (e) a data collection system adapted to receive input from the detector.

36. The apparatus of claim 35, wherein the nucleic acids have known sequences.

37. A method of analyzing a receptor bound to an array of polymers, comprising:

122

providing an array of different polymers immobilized on a substrate, the different polymers occupying different localized areas on the substrate having an area of less than 0.01 cm², at least some of the nucleic acids being bound to a labeled receptor; and
 detecting the localized areas of the array with a detector, each localized area being scanned for one second or less, to detect which polymers bind to the labeled receptor.

38. The method of claim 37, wherein the different polymers are different nucleic acids.

39. The method of claim 37, wherein the receptor is a nucleic acid.

40. The method of claim 37, wherein the different polymers are peptides.

41. A method for screening comprising:
 exposing a substrate having a ligand in a localized area to a marked receptor; and
 detecting binding between the ligand and the marked receptor by scanning the substrate for 1 second or less in the localized area.

42. The method of claim 41, wherein the ligand is a nucleic acid.

43. The method of claim 41, wherein the marked receptor is a marked nucleic acid.

44. The method of claim 41, wherein the ligand is a polypeptide.

45. The method of claim 41, wherein the receptor is marked with a fluorescent label.

46. The method of claim 41, wherein the substrate has more than 1000 different ligands in 1000 different localized areas within an area less than 1 cm².

47. The method of claim 41, wherein the receptor is marked with a fluorescent marker.

48. The apparatus of claim 35, wherein the ten different nucleic acids have lengths up to 100 nucleotides.

49. An apparatus for detection of fluorescent marked locations on a substrate, wherein the substrate comprises a plurality of different polymer sequences coupled to a surface of the substrate and wherein the plurality of different polymer sequences comprises a plurality of different nucleic acid sequences, each of the different polymer sequences being coupled in a different localized area of the surface at a density of more than 1000/cm², each of the localized areas having an area of 10⁻³ cm² or less, the apparatus comprising:
 (a) a light source constructed and arranged to direct light at a plurality of the localized areas;
 (b) a detector constructed and arranged to detect light fluoresced from the plurality of the localized areas in response to the light source;
 (c) a translator constructed and arranged to translate the substrate relative to the light source whereby the plurality of the localized areas are exposed to the light; and
 (d) a data storage system constructed and arranged to store fluoresced light intensity as a function of location on the surface, the data storage system coupled to the translator and the detector.

50. An apparatus for detection of fluorescent markers at a plurality of localized areas on a substrate, comprising:
 (a) a light source constructed and arranged to direct light at the plurality of localized areas;
 (b) a detector constructed and arranged to detect light fluoresced from the plurality of localized areas in response to the light source;
 (c) a translator constructed and arranged to translate the substrate relative to the light source whereby the plurality of localized areas are exposed to the light; and

US 6,403,320 B1

123

(d) a data storage system constructed and arranged to store fluoresced light intensity as a function of location on the surface, the data storage system coupled to the translator and the detector;

wherein the substrate comprises a plurality of different polymer sequences coupled to a surface of the substrate and wherein the plurality of different polymer sequences comprises a plurality of different nucleic acid sequences, each of the different polymer sequences being coupled in a different of the localized areas, each localized area having an area of 10^{-2} cm² or less.

51. An apparatus for detection of fluorescent markers at a plurality of localized areas on a substrate, comprising:

(a) a light source constructed and arranged to direct light at the plurality of localized areas, wherein the substrate comprises a plurality of different polymer sequences coupled to a surface of the substrate and wherein the plurality of different polymer sequences comprises a plurality of different nucleic acid sequences, each of the different polymer sequences being coupled in a different of the localized areas, each of the localized areas having an area of 10^{-2} cm² or less;

(b) a detector constructed and arranged to detect light fluoresced from the plurality of localized areas in response to the light source;

(c) a translator constructed and arranged to translate the substrate relative to the light source whereby the plurality of localized areas are exposed to the light; and

(d) a data storage system constructed and arranged to store fluoresced light intensity as a function of location on the surface, the data storage system coupled to the translator and the detector.

52. A method in accordance with claim 1 further comprising collecting and analyzing data from the detection step using a computer system.

53. A method in accordance with claim 52 wherein the computer system comprises an IBM PC or AT compatible computer.

54. A method in accordance with claim 1 further comprising translating the array using an X/Y translation table.

55. A method in accordance with claim 54 wherein the X/Y translation table is controlled by a digital computer.

56. A method in accordance with claim 1 wherein the detecting step uses a photomultiplier tube.

57. A method in accordance with claim 1 wherein the detecting step uses a pre-amplifier.

58. A method in accordance with claim 1 further comprising displaying data that has been acquired using a video display.

59. A method in accordance with claim 1 further comprising detecting and displaying the output of the detection using a photomultiplier tube, a pre-amplifier, a computer, and a video display.

60. A method in accordance with claim 1 wherein the detecting step involves counting the number of photons per unit area.

61. A method in accordance with claim 1 wherein the detecting step includes acquiring data every 1 to 100 μ m.

62. A method in accordance with claim 1 wherein the detecting step includes acquiring data with a data collection diameter of about 0.8 to 10 μ m.

63. A method in accordance with claim 1 further comprising exposing the array to laser light.

64. A method in accordance with claim 1 further comprising exposing the array to laser light having a wavelength of greater than 520 nm.

65. A method in accordance with claim 1 further comprising translating the array with an X/Y translator under the

124

control of a computer; detecting and displaying the output of the detection using a photomultiplier tube, a pre-amplifier, and a computer; and collecting and analyzing data from the detection step using a computer system.

66. A method in accordance with claim 37 further comprising collecting and analyzing data from the detection step using a computer system.

67. A method in accordance with claim 66 wherein the computer system comprises an IBM PC or AT compatible computer.

68. A method in accordance with claim 37 further comprising translating the array using an X/Y translation table.

69. A method in accordance with claim 37 wherein the X/Y translation table is controlled by a digital computer.

70. A method in accordance with claim 37 wherein the detecting step uses a photomultiplier tube.

71. A method in accordance with claim 37 wherein the detecting step uses a pre-amplifier.

72. A method in accordance with claim 37 further comprising displaying data that has been acquired using a video display.

73. A method in accordance with claim 37 further comprising detecting and displaying the output of the detection using a photomultiplier tube, a pre-amplifier, a computer, and a video display.

74. A method in accordance with claim 37 wherein the detecting step involves counting the number of photons per unit area.

75. A method in accordance with claim 37 wherein the detecting step includes acquiring data every 1 to 100 μ m.

76. A method in accordance with claim 37 wherein the detection step includes acquiring data with a data collection diameter of about 0.8 to 10 μ m.

77. A method in accordance with claim 37 further comprising exposing the array to laser light.

78. A method in accordance with claim 37 further comprising exposing the array to laser light having a wavelength of greater than 520 nm.

79. A method in accordance with claim 37 further comprising translating the array with an X/Y translator under the control of a computer; detecting and displaying the output of the detection using a photomultiplier tube, a pre-amplifier, and a computer; and collecting and analyzing data from the detection step using a computer system.

80. An apparatus in accordance with claim 37 further comprising a computer system.

81. An apparatus in accordance with claim 80 wherein the computer system comprises an IBM PC or AT compatible computer.

82. An apparatus in accordance with claim 25 further comprising an X/Y translation table.

83. An apparatus in accordance with claim 25 wherein the X/Y translation table is controlled by a digital computer.

84. An apparatus in accordance with claim 25 further comprising a photomultiplier tube.

85. An apparatus in accordance with claim 25 further comprising a pre-amplifier.

86. An apparatus in accordance with claim 25 further comprising a video display.

87. An apparatus in accordance with claim 25 further comprising a photomultiplier tube, a pre-amplifier, a computer, and a video display.

88. An apparatus in accordance with claim 29 further comprising a laser.

89. An apparatus in accordance with claim 29 further comprising an X/Y translator under the control of a computer; a photomultiplier tube, a pre-amplifier, and a computer; and a computer system.

US 6,403,320 B1

125

90. An apparatus in accordance with claim 29 further comprising a computer system.

91. An apparatus in accordance with claim 90 wherein the computer system comprises an IBM PC or AT compatible computer.

92. An apparatus in accordance with claim 37 further comprising an X/Y translation table.

93. An apparatus in accordance with claim 92 wherein the X/Y translation table is controlled by a digital computer.

94. An apparatus in accordance with claim 29 further comprising a photomultiplier tube.

95. An apparatus in accordance with claim 29 further comprising a pre-amplifier.

96. An apparatus in accordance with claim 29 further comprising a video display.

97. An apparatus in accordance with claim 29 further comprising a photomultiplier tube, a pre-amplifier, a computer, and a video display.

98. An apparatus in accordance with claim 29 further comprising a laser.

99. An apparatus in accordance with claim 29 further comprising an X/Y translator under the control of a computer; a photomultiplier tube, a pre-amplifier, and a computer; and a computer system.

100. A method in accordance with claim 23 wherein the receptor is a nucleic acid.

101. An apparatus in accordance with claim 25 wherein the receptor is a nucleic acid.

102. An apparatus in accordance with claim 29 wherein the receptor is a nucleic acid.

103. A method in accordance with claim 25 wherein the receptor is a peptide.

104. A method in accordance with claim 37 wherein the receptor is a peptide.

105. A method in accordance with claim 29 wherein the receptor is a peptide.

106. A method for detecting binding between a nucleic acid probe bound to a substrate and a nucleic acid target comprising:

providing an array of positionally distinguishable nucleic acids of known sequence bound to a substrate having a density of more than 300 different nucleic acids/cm²; and

detecting a double stranded nucleic acid.

107. A method in accordance with claim 106 wherein the density of the array is 1,000 different nucleic acids/cm².

108. A method in accordance with claim 107 wherein the density of the array is more than 10,000 different nucleic acids/cm².

109. A method of analyzing using an array of nucleic acids, comprising:

providing an array of different nucleic acids of known sequence immobilized on a substrate at a density more than 1,000 nucleic acids/cm², the different nucleic acids being positionally distinguishable and, at least some of the nucleic acids being bound to a fluorescently labeled receptor;

illuminating an area of the array greater than 10⁻³ cm² with an energy source; and

detecting fluoresced light from the known locations of the array with a detector, the detection step being one second or less, to detect which nucleic acids bind to the labeled receptor.

110. A method in accordance with claim 109 wherein the receptor is a nucleic acid.

126

111. A method in accordance with claim 109 wherein the receptor is a peptide.

112. A method in accordance with claim 109 wherein the density is more than 10,000 nucleic acids/cm².

113. A method in accordance with claim 1 wherein the sequences of the nucleic acids are known.

114. An apparatus in accordance with claim 41 wherein the sequence of the ligand is known.

115. An apparatus in accordance with claim 49 wherein the sequences of the nucleic acids are known.

116. An apparatus in accordance with claim 50 wherein the sequences of the nucleic acids are known.

117. An apparatus in accordance with claim 51 wherein the sequences of the nucleic acids are known.

118. An apparatus in accordance with claim 35 wherein the density of the different nucleic acid sequences on the array is more than 1000/cm².

119. An apparatus in accordance with claim 37 wherein the density of the different nucleic acid sequences on the array is more than 1000/cm².

120. An apparatus in accordance with claim 41 wherein the density of the different nucleic acid sequences on the array is more than 1000/cm².

121. An apparatus in accordance with claim 50 wherein the density of the different nucleic acid sequences on the array is more than 1000/cm².

122. An apparatus in accordance with claim 51 wherein the density of the different nucleic acid sequences on the array is more than 1000/cm².

123. An apparatus in accordance with claim 1 wherein the density of the different nucleic acid sequences on the array is more than 300/cm².

124. An apparatus in accordance with claim 1 wherein the density of the different nucleic acid sequences on the array is more than 300/cm².

125. An apparatus in accordance with claim 35 wherein the density of the different nucleic acid sequences on the array is more than 300/cm².

126. An apparatus in accordance with claim 29 wherein the density of the different nucleic acid sequences on the array is more than 300/cm².

127. A method of analyzing using an array of polymers, comprising:

providing an array of different polymers of known sequence immobilized on a substrate at a density more than 1,000 polymers/cm², the different polymers being positionally distinguishable and at least some of the polymers being bound to a fluorescently labeled target polymers;

illuminating an area of the array greater than 10⁻³ cm² with an energy source; and

detecting fluoresced light from the known locations of the array with a detector, the detection step being one second or less, to detect which polymers on the array bind to the labeled target polymers.

128. A method in accordance with claim 127 wherein the polymers are nucleic acids.

129. A method in accordance with claim 127 wherein the polymers are peptides.

130. A method in accordance with claim 128 the density of the different nucleic acids is more than 10,000/cm².

131. A method in accordance with claim 129 wherein the density of the different peptides is more than 10,000/cm².

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 6,403,320 B1
DATED : June 11, 2002
INVENTOR(S) : J. Leighton Read et al.

Page 1 of 10

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 1,

Lines 8-9, delete “now U.S. Pat. No. 6,197,506”;
Lines 12, 14, 17 and 28, delete “now abandoned”; and
Line 66, after “underlying” delete “the”.

Column 2,

Line 8, after “sequence” insert -- of --; and
Line 60, after “VLSIPS” insert -- TM Technology (Very Large Scale Immobilized Polymer Synthesis) --.

Column 4,

Line 31, after “probes” delete “are”.

Column 5,

Line 5, delete “upon”.

Column 6,

Line 48, delete “peptide” and insert therefor -- TM Technology nucleotide --;
Line 50, delete “dipeptide” and insert therefor -- TM Technology --; and
Line 52, after “VLSIPS” add -- TM Technology --.

Column 10,

Lines 8, 21 and 30, after “VLSIPS” add -- TM Technology --.

Column 12,

Line 6, after “VLSIPS” add -- TM Technology --;
Line 10, delete “(VLSIPS) technology” and insert therefor -- (VLSIPS TM) Technology --;
Line 23, delete “technology” and insert therefor -- TM Technology --;
Line 30, delete “U.S. Ser. No. 07/492,462 (VLSIPS CIP)” and insert therefor -- Pirrung et al. (1992) U.S. Pat. No. 5,143,854 --;
Line 37, after “VLSIPS” add -- TM Technology --;
Lines 38-39, delete “U.S. Ser. No. 07/492,462 (VLSIPS CIP)” and insert therefor -- Pirrung et al. (1992) U.S. Pat. No. 5,143,854 --;
Lines 39-40, delete “U.S. Ser. No. 07/624,120 (automated VLSIPS)” and insert therefor -- U.S. Pat. No. 5,489,678 --;
Line 41, after “VLSIPS” add -- TM Technology --;
Line 47, delete “technology” and insert therefor -- TM Technology --;
Line 51, after “10²” add -- /cm² --; and
Line 62, delete “U.S. Ser. No. 07/624,1200 (automated VLSIPS) and U.S. Ser. No. 07/626,730, (sequencing by synthesis)” and insert therefor -- U.S. Pat. No. 5,489,678 and U.S. Pat. No. 5,427,408 --.

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 6,403,320 B1
DATED : June 11, 2002
INVENTOR(S) : J. Leighton Read et al.

Page 2 of 10

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 13,

Line 23, after "VLSIPS" add -- TM Technology --.

Column 45,

Lines 36-37, delete "U.S. Ser. No. 07/492,462 (VLSIPS CIP)" and insert therefor -- Pirrung et al. (1992) U.S. Pat. No. 5,143,854 --; and

Line 40, delete "U.S. Ser. No. 07/624,120 (automated VLSIPS)" and insert therefor -- U.S. Pat. No. 5,489,678 --.

Column 47,

Line 17, after "VLSIPS" add -- TM Technology --;

Line 52, delete "(VLSIPS parent)" and insert therefor -- from which CIP --;

Line 53, delete "(VLSIPS CIP,)" and insert therefor -- issued as U.S. Pat. No. 5,143,854, and --;

Lines 53-54, delete "(caged biotin parent), and" and insert therefor -- from which CIP --; and

Lines 54-55, delete "(caged biotin CIP)" and insert therefor -- issued as U.S. Pat. No. 5,252,743 --.

Column 48,

Line 9, delete "(VLSIPS parent); or" and insert therefor -- from which CIP --;

Line 10, delete "(VLSIPS CIP)" and insert therefor -- issued as U.S. Pat. No. 5,143,854 --;

Lines 10-11, delete "U.S. Ser. No. 07/624,120, (automated VLSIPS)" and insert therefor -- U.S. Pat. No. 5,489,678 --;

Lines 20-21, delete "U.S. Ser. No. 07/624,120, (automated VLSIPS)" and insert therefor -- U.S. Pat. No. 5,489,678 --;

Line 57, delete "Theoretical" and insert therefor -- Theoretical --; and

Line 60, delete "technology" and insert therefor -- TM Technology --.

Column 49,

Line 2, delete "e.e." and insert therefor -- e.g. --;

Line 4, delete "Khropko" and insert therefor -- Khrapko --; and

Lines 29 and 43, after "VLSIPS" add -- TM Technology --.

Column 50,

Line 10, after "that" add -- the --.

Column 51,

Line 29, after "way" delete -- , --.

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 6,403,320 B1
DATED : June 11, 2002
INVENTOR(S) : J. Leighton Read et al.

Page 3 of 10

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 54,

Line 21, delete "U.S. Ser. No. 07/624,120 (" and insert therefor
-- U.S. Pat. No. 5,489,678 --; and after "VLSIPS" delete ")" and insert therefor
-- TM Technology --.

Column 55,

Line 17, after "VLSIPS" add -- TM Technology --;
Lines 19-20, delete "U.S. Ser. No. 07/435,316 (caged biotin parent), and U.S. Ser. No. 07/612,671 (caged biotin CIP)" and insert therefor -- Barrett et al. (1993) U.S. Pat. No. 5,252,743 --;
Line 28, delete "VISIPS" and insert therefor -- VLSIPSTM Technology --;
Lines 31, 32 and 35, after "VLSIPS" add -- TM Technology --;
Line 58, delete "affixed" and insert therefor -- affixing them --; and
Lines 60-61, delete "U.S. Ser. No. 07/362,901 (VLSIPS parent); U.S. Ser. No. 07/492,462 (VLSIPS CIP)" and insert therefor -- Pirrung et al. (1992) U.S. Pat. No. 5,143,854 --.

Column 56,

Line 19, after "particularly" add -- through the --;
Lines 24-26, delete "U.S. Ser. No. 07/492,462 (VLSIPS CIP); U.S. Ser. No. 07/624,120 (automated VLSIPS); and U.S. Ser. No. 07/612,671 (caged biotin CIP)" and insert therefor -- Pirrung et al. (1992) U.S. Pat. No. 5,143,854; U.S. Pat. No. 5,489,678; and Barrett et al. (1993) U.S. Pat. No. 5,252,743 --;
Lines 31-32, delete "U.S. Ser. No. 07/492,462 (VLSIPS CIP); and U.S. Ser. No. 07/624,120, (automated VLSIPS)" and insert therefor -- Pirrung et al. (1992) U.S. Pat. No. 5,143,854; and U.S. Pat. No. 5,489,678 --;
Lines 45-46, delete "U.S. Ser. No. 07/492,462 (VLSIPS CIP) and U.S. Ser. No. 07/624,120, (automated VLSIPS)" and insert therefor -- U.S. Pat. No. 5,489,678 --; and
Line 61, delete "are" and insert -- is --.

Column 57,

Lines 6-8, delete "U.S. Ser. No. 07/624,120 (automated VLSIPS) which was filed simultaneously with this application" and insert therefor -- U.S. Pat. No. 5,489,678 --;
Lines 17-18, delete "U.S. Ser. No. 07/624,120, (automated VLSIPS)" and insert therefor -- U.S. Pat. No. 5,489,678 --;
Line 34, delete "converse" and insert therefor -- confer --; and
Line 39, delete "technology" and insert therefor -- TM Technology --.

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 6,403,320 B1
DATED : June 11, 2002
INVENTOR(S) : J. Leighton Read et al.

Page 4 of 10

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 58,

Line 10, after “to” add -- the --;
Line 48, after “that” delete -- it --;
Line 48, after “will” add -- allow the probe to --; and
Line 51, delete “close” and insert therefor -- closely --.

Column 60,

Lines 5, 12 and 32, delete “arylammonium” and insert therefor -- alkylammonium --;
Line 33, after “manner” add -- such --;
Line 39, after “specific” add -- such --;
Line 43, after “VLSIPS” add -- TM Technology --;
Lines 46-47, delete “VLZIPS” and insert therefor -- VLSIPSTM Technology --; and
Lines 58-60, delete “U.S. Ser. No. 07/492,462 (VLSIPS CIP) and U.S. Ser. No. 07/624,120 (automated VLSIPS)” and insert therefor -- Pirrung et al. (1992) U.S. Pat. No. 5,143,854 and U.S. Pat. No. 5,489,678 --.

Column 61,

Line 60, after “VLSIPS” add -- TM Technology --;
Line 61, delete “technology”; and
Line 63, delete “technology” and insert therefor -- TM Technology --.

Column 62,

Lines 15, 17, 23, 35, 54, 56, 57, 65 and 67, after “VLSIPS” add -- TM Technology --.

Column 63,

Line 2, after “VLSIPS” add -- TM Technology --;
Line 46, delete “U.S. Ser. No. 07/612,671 (caged biotin CIP)” and insert therefor -- U.S. Pat. No. 5,252,743 --; and
Line 52, after “VLSIPS” add -- TM Technology --.

Column 64,

Line 5, after “VLSIPS” add -- TM Technology --; and
Line 9, delete “U.S. Ser. No. 07/626,730, (sequencing by synthesis);” and insert therefor -- U.S. Pat. No. 5,547,839 --.

Column 65,

Lines 4, 20 and 61, after “VLSIPS” add -- TM Technology --;
Line 28, after “complex” delete “, it” and insert therefor -- . It --; and
Line 64, delete “expressed”.

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 6,403,320 B1
DATED : June 11, 2002
INVENTOR(S) : J. Leighton Read et al.

Page 5 of 10

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 66,

Line 52, after "VLSIPS" add -- TM Technology --; and delete "use" and insert -- used --.

Column 67,

Line 35, after "susceptibilities," add -- and --.

Column 68,

Line 9, delete "fingerprinted" and insert therefor -- fingerprinting --; and delete "in" and insert -- an --; and

Line 10, delete "mosaism" and insert therefor -- mosaicism --.

Column 69,

Lines 11-12, delete "Again, the target sequences may be desired to be fragmented" and insert therefor -- Again, it may be desirable to fragment the target sequences --.

Column 70,

Line 31, after "VLSIPS" add -- TM Technology --; and

Line 46, after "allows" delete "the".

Column 71,

Line 29, delete "to define" and insert therefor -- in defining --;

Line 40, delete "as" and insert therefor -- are --; and

Line 65, delete "upon" and insert therefor -- by --.

Column 72,

Line 19, delete "amendable" and insert therefor -- detectable --;

Line 28, delete "present" and insert therefor -- presently available --;

Line 29, delete "screen" and insert therefor -- screens --;

Line 44, delete "also is" and insert therefor -- is also --;

Lines 59-61, delete "U.S. Ser. No. 07/492,462 (VLSIPS CIP), and U.S. Ser. No. 07/624,120, (automated VLSIPS)" and insert therefor -- Pirrung et al. (1992) U.S. Pat. No. 5,143,854; and U.S. Pat. No. 5,489,678 --; and

Line 67, after "VLSIPS" add -- TM Technology --.

Column 73,

Line 1, after "VLSIPS" add -- TM Technology --; and

Lines 6-7, delete "U.S. Ser. No. 07/462,492 (VLSIPS CIP) and U.S. Ser. No. 07/624,120, (automated VLSIPS)" and insert therefor -- Pirrung et al. (1992) U.S. Pat. No. 5,143,854; and U.S. Pat. No. 5,489,678 --.

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 6,403,320 B1
DATED : June 11, 2002
INVENTOR(S) : J. Leighton Read et al.

Page 6 of 10

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 74,

Lines 6-7, delete “ U.S. Ser. No. 07/624,120, (automated VLSIPS)” and insert therefor -- U.S. Pat. No. 5,489,678 --;
Line 24, delete “then” and insert therefor -- them --;
Lines 36, 43 and 47, after “VLSIPS” add -- TM Technology --;
Lines 48-49, delete “U.S. Ser. No. 07/492,462 (VLSIPS CIP) and U.S. Ser. No. 07/624,120, (automated VLSIPS)” and insert therefor -- Pirrung et al. (1992) U.S. Pat. No. 5,143,854 and U.S. Pat. No. 5,489,678 --;
Line 53, delete “ U.S. Ser. No. 07/624.120 (automated VLSIPS)” and insert therefor -- U.S. Pat. No. 5,489,678 --; and
Line 66, delete “U.S. Ser. No. 07/492,462” and insert therefor -- which issued as U.S. Pat. No. 5,143,854 --.

Column 75,

Line 11, delete “U.S. Ser. No. 07/492,462 (VLSIPS CIP) and U.S. Ser. No. 07/624,120 (automated VLSIPS)” and insert therefor -- Pirrung et al. (1992) U.S. Pat. No. 5,143,854 and U.S. Pat. No. 5,489,678 --;
Line 16, after “VLSIPS” add -- TM Technology --; and
Line 66, delete “anaines” and insert therefor -- amines --.

Column 78,

Line 18, after “VLSIPS” add -- TM Technology --; and delete “have been” and insert therefor -- include, for example --.

Column 81,

Line 34, after “functionalized” add -- , --; and
Line 35, after “(see example below)” add -- : --.

Column 82,

Line 1, delete “group” and insert therefor -- groups --;
Line 2, after “acetate” add -- : --;
Line 24, after “ether” delete “.” and insert therefor -- : --; and
Line 54, delete “Related to” and insert therefor -- With respect to --;
Line 62, after “VLSIPS” add -- TM Technology --; and
Line 63, delete “U.S. Ser. No. 07/624,120, (automated VLSIPS)” and insert therefor -- U.S. Pat. No. 5,489,678 --.

Column 84,

Line 6, after “VLSIPS” add -- TM Technology --;
Line 44, delete “in our labs”; and
Line 49, delete “Brosystems” and insert therefor -- Biosystems --.

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 6,403,320 B1
DATED : June 11, 2002
INVENTOR(S) : J. Leighton Read et al.

Page 7 of 10

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 85,

Line 41, delete “U.S. Ser. No. 07/612,671 (caged biotin CIP)” and insert -- Barrett et al. (1993) U.S. Pat. No. 5,252,743 --.

Column 86,

Lines 13 and 14, after “VLSIPS” add -- TM Technology --.

Column 87,

Line 24, delete “Cell” and insert therefor -- In addition to cell --;
Line 41, after “VLSIPS” add -- TM --;
Line 44, delete “CARIOS” and insert therefor -- CABIOS --; and
Line 49, after “VLSIPS” add -- TM Technology --.

Column 89,

Line 2, after “*coli*” delete “)”.

Column 90,

Line 13, delete “ficoerythrin” and insert therefor -- phycoerythrin --; and
Line 15, delete “primary” and insert therefor -- primarily --.

Column 91,

Line 47, delete “U.S. Ser. No. 07/649,642 (VLSIPS CIP); and U.S. Ser. No. 07/624,120, (automated VLSIPS)” and insert therefor -- Pirrung et al. (1992) U.S. Pat. No. 5,143,854 and U.S. Pat. No. 5,489,678 --; and
Line 56, delete “U.S. Ser. No. 07/624,120, (VLSIPS)” and insert therefor -- U.S. Pat. No. 5,489,678 --.

Column 92,

Lines 23-24, delete “U.S. Ser. No. 07/624,120, (automated VLSIPS)” and insert therefor -- U.S. Pat. No. 5,489,678 --;
Line 39, after “of” add -- a --; and
Line 47, after “fragmented” delete “up”.

Column 94,

Lines 45, 52, 60 and 67, after “VLSIPS” add -- TM Technology --; and
Line 47, delete “simply” and insert therefor -- simplify --.

Column 95,

Line 2, after “adjacent” add -- to --;
Lines 5 and 33, after “VLSIPS” add -- TM Technology --;
Line 44, after “utilized” add -- for --; and
Line 66, after “sequencing” delete “of”.

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 6,403,320 B1
DATED : June 11, 2002
INVENTOR(S) : J. Leighton Read et al.

Page 8 of 10

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 96,

Line 58, after "VLSIPS" add -- TM --; and

Lines 60-61, delete "U.S. Ser. No. 07/612,671 (caged biotin CIP)" and insert therefor -- Barrett et al. (1993) U.S. Pat. No. 5,252,743 --.

Column 97,

Line 26, after "VLSIPS" add -- TM Technology --; and

Line 56, delete "An" and insert therefor -- In --.

Column 98,

Line 26, delete "preventative" and insert -- preventive --.

Column 99,

Line 29, delete "(VLSIPS parent), filed Jun. 7, 1989" and insert therefor -- , from which CIP 07/492,462 issued as U.S. Pat. No. 5,143,854 --;

Lines 29-30, delete "Pirrung et al., U.S. Ser. No. 07/492,462 (VLSIPS CIP), filed Mar. 7, 1990; Barrett et al."; and

Lines 33-36, delete "(caged biotin CIP), filed Nov. 13, 1990; and commonly assigned and simultaneously filed applications U.S. Ser. No. 07/624,120, (automated VLSIPS) and U.S. Ser. No. 07/626,730, (sequencing by synthesis)" and insert therefor -- from which CIP 07/612,671 issued as U.S. Pat. No. 5,252,743 and U.S. Pat. No. 5,489,678 --.

Column 100,

Lines 29-30, delete "5'-O-nitrovertryl" and insert therefor -- 5'-O-nitroveratryl --; and

Line 31, delete "5'-O-nitrovertryl" and insert therefor -- 5'-O-nitroveratryl --.

Column 101,

Line 21, after "VLSIPS" add -- TM Technology --.

Column 103,

Line 48, after "added" add -- to --; and

Line 56, after "was" delete "a".

Column 106,

Lines 30-31, delete "U.S. Ser. No. 07/492,462 (VLSIPS CIP) and U.S. Ser. No. 07/624,120, (automated VLSIPS)" and insert therefor -- Pirrung et al. (1992) U.S. Pat. No. 5,143,854 and U.S. Pat. No. 5,489,678 --; and

Lines 52-54, delete "U.S. Ser. No. 07/492,462 (VLSIPS CIP), and U.S. Ser. No. 07/624,120, (automated VLSIPS)" and insert therefor -- Pirrung et al. (1992) U.S. Pat. No. 5,143,854 and U.S. Pat. No. 5,489,678 --.

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 6,403,320 B1
DATED : June 11, 2002
INVENTOR(S) : J. Leighton Read et al.

Page 9 of 10

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 107.

Lines 56-58, delete “U.S. Ser. No. 07/492,462 (VLSIPS CIP); and U.S. Ser. No. 07/624,120” and insert therefor -- Pirrung et al. (1992) U.S. Pat. No. 5,143,854 --;

Column 108.

Lines 16-17, delete “U.S. Ser. No. 07/492,462 (VLSIPS CIP)” and insert therefor -- Pirrung et al. (1992) U.S. Pat. No. 5,143,854 --;

Lines 27-28, delete “U.S. Ser. No. 07/624,120” and insert therefor -- U.S. Pat. No. 5,489,678 --;

Line 39, delete “U.S. Ser. No. 07/492,462 (VLSIPS CIP)” and insert -- Pirrung et al. (1992) U.S. Pat. No. 5,143,854 --;

Line 48, delete “durapore” and insert -- DuraporeTM --; and

Line 49, after “Durapore” add -- TM --.

Column 109.

Lines 26-27, delete “U.S. Ser. No. 07/492,462 (VLSIPS CIP)” and insert therefor -- Pirrung et al. (1992) U.S. Pat. No. 5,143,854 --; and

Line 62, delete “U.S. Ser. No. 07/492,462 (VLSIPS CIP)” and insert therefor -- Pirrung et al. (1992) U.S. Pat. No. 5,143,854 --.

Column 110.

Line 3, delete “NOVC” and insert therefor -- NVOC --;

Line 11, delete “U.S. Ser. No. 07/492,462 (VLSIPS CIP)” and insert therefor -- Pirrung et al. (1992) U.S. Pat. No. 5,143,854 --; and

Line 67, delete “that”.

Column 111.

Line 26, delete “U.S. Ser. No. 07/492,462 (VLSIPS CIP)” and insert therefor -- Pirrung et al. (1992) U.S. Pat. No. 5,143,854 --.

Column 113.

Line 2, delete “(caged biotin parent) and” and insert therefor -- from which CIP --;

Lines 2-3, delete “(caged biotin CIP)” and insert therefor -- issued as U.S. Pat. No. 5,252,743 --; and

Line 52, delete “are” and insert -- is --.

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 6,403,320 B1
DATED : June 11, 2002
INVENTOR(S) : J. Leighton Read et al.

Page 10 of 10

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 114.

Lines 23-24, delete "U.S. Ser. No. 07/612,671 (caged biotin CIP)" and insert therefor -- Barrett et al. (1993) U.S. Pat. No. 5,252,743 --; and
Line 33, after "VLSIPS" add -- TM Technology --.

Column 116.

Line 39, before "non-hybridization" add -- and --.

Column 117.

Line 41, after "environment" add -- to --.

Column 118.

Line 14, delete "her" and insert therefor -- their --.

Column 119.

Lines 11, 19, 34 and 36, after "VLSIPS" add -- TM Technology --.

Column 122.


Line 4, delete "nucleic acids" and insert -- different polymers --.

Column 126.

Line 18, delete "apparatus" and insert therefor -- method --;
Line 18, delete "37" and insert therefor -- 38 --; and
Line 61, after "**128**" add -- wherein --.

Signed and Sealed this

Second Day of December, 2003

A handwritten signature in black ink, appearing to read "James E. Rogan", with a long horizontal flourish extending from the bottom of the signature.

JAMES E. ROGAN
Director of the United States Patent and Trademark Office

EXHIBIT 3

US006420169B1

(12) **United States Patent**
Read et al.(10) **Patent No.: US 6,420,169 B1**
(45) **Date of Patent: *Jul. 16, 2002**(54) **APPARATUS FOR FORMING
POLYNUCLEOTIDES OR POLYPEPTIDES**(75) Inventors: **J. Leighton Read; Stephen P.A.
Fodor**, both of Palo Alto; **Lubert
Stryer**, Stanford, all of CA (US);
Michael C. Pirrung, Mesquite, TX
(US); **Paul D. Hoeprich, Jr.**, Danville,
CA (US)(73) Assignee: **Affymetrix, Inc.**, Santa Clara, CA (US)(*) Notice: This patent issued on a continued pro-
secution application filed under 37 CFR
1.53(d), and is subject to the twenty year
patent term provisions of 35 U.S.C.
154(a)(2).Subject to any disclaimer, the term of this
patent is extended or adjusted under 35
U.S.C. 154(b) by 0 days.(21) Appl. No.: **08/348,471**(22) Filed: **Nov. 30, 1994****Related U.S. Application Data**(63) Continuation of application No. 07/805,727, filed on Dec. 6,
1991, now Pat. No. 5,424,186, which is a continuation-in-
part of application No. 07/492,462, filed on Mar. 7, 1990,
now Pat. No. 5,143,854, which is a continuation-in-part of
application No. 07/362,901, filed on Jun. 7, 1989, now
abandoned, said application No. 07/805,727, is a continua-
tion-in-part of application No. 07/624,120, filed on Dec. 6,
1990, now abandoned, which is a continuation-in-part of
application No. 07/492,462, which is a continuation-in-part
of application No. 07/362,901.(51) **Int. Cl.**⁷ **C12M 1/36; G01N 33/543;**
A61K 38/00(52) **U.S. Cl.** **435/289.1; 435/DIG. 43;**
435/DIG. 44; 422/131; 436/518; 935/88;
530/335(58) **Field of Search** **435/6, 969, 289.1,**
435/292.1, DIG. 43, DIG. 44; 536/25.3;
436/518, 527, 807; 530/335; 422/131; 935/88(56) **References Cited****U.S. PATENT DOCUMENTS**

3,849,137 A	11/1974	Barzynski et al.
3,862,056 A	1/1975	Hartman
3,862,932 A *	1/1975	Sellsted et al. 260/239.1
3,939,350 A	2/1976	Arwin et al.
4,072,576 A	2/1978	Arwin et al.
4,180,739 A	12/1979	Abu-Shumays
4,238,757 A	12/1980	Schenck
4,269,933 A	5/1981	Pazos
4,314,821 A	2/1982	Rice
4,327,073 A	4/1982	Huang
4,339,528 A	7/1982	Goldman
4,342,905 A	8/1982	Fujii et al.
4,373,071 A	2/1983	Itakura
4,405,771 A	9/1983	Jagur
4,444,878 A	4/1984	Paulus
4,444,892 A	4/1984	Malmros
4,448,534 A	5/1984	Wert et al.

4,458,066 A	7/1984	Caruthers et al.
4,483,920 A	11/1984	Gillespie et al.
4,500,707 A	2/1985	Caruthers et al.
4,500,919 A	2/1985	Schreiber
4,516,833 A	5/1985	Fusek
4,517,338 A *	5/1985	Urdea et al. 536/25.3
4,537,861 A	8/1985	Elings et al.
4,542,102 A	9/1985	Dattagupta et al.
4,555,490 A	11/1985	Merril
4,562,157 A	12/1985	Lowe et al.
4,569,967 A	2/1986	Kornreich et al.
4,580,895 A	4/1986	Patel
4,584,277 A	4/1986	Ullman
4,613,566 A	9/1986	Potter
4,624,915 A	11/1986	Schindler et al.
4,626,684 A	12/1986	Landa
4,631,211 A	12/1986	Houghten
4,637,861 A	1/1987	Krull et al.
4,677,054 A	6/1987	White et al.
4,681,859 A	7/1987	Kramer
4,683,202 A	7/1987	Mullis
4,689,405 A	8/1987	Frank et al.
4,704,353 A	11/1987	Humphries et al.
4,711,955 A	12/1987	Ward et al.
4,713,326 A	12/1987	Dattagupta et al.
4,713,347 A	12/1987	Mitchell et al.
4,719,179 A	1/1988	Barany
4,719,615 A	1/1988	Feyrer et al.
4,722,906 A	2/1988	Guire
4,728,502 A	3/1988	Hamill
4,728,591 A	3/1988	Clark et al.
4,731,325 A	3/1988	Palva et al.
4,755,458 A	7/1988	Rabbani et al.
4,762,881 A	8/1988	Kauer
4,777,019 A	10/1988	Dandekar

(List continued on next page.)

FOREIGN PATENT DOCUMENTS

DE	2242394	3/1974
DE	3440141	5/1986
DE	3505287	3/1988

(List continued on next page.)

OTHER PUBLICATIONSDi Mauro and Hollenberg, "DNA Technology in Chip Con-
struction," *Adv. Mater.*, 5:384-386 (1993).

(List continued on next page.)

Primary Examiner—Padmashri Ponnaluri(74) *Attorney, Agent, or Firm*—Townsend and Townsend
and Crew LLP(57) **ABSTRACT**

A method for synthesizing oligonucleotides on a solid substrate. The method provides for the irradiation of a first predefined region of the substrate without irradiation of a second predefined region of the substrate. The irradiation step removes a protecting group therefrom. The substrate is contacted with a first nucleotide to couple the nucleotide to the substrate in the first predefined region. By repeating these steps, an array of diverse oligonucleotides is formed on the substrate.

15 Claims, 42 Drawing SheetsMicrofiche Appendix Included
(5 Microfiche, 348 Pages)

US 6,420,169 B1

Page 2

U.S. PATENT DOCUMENTS

4,780,504 A	10/1988	Buendia et al.	5,445,934 A	8/1995	Fodor et al.
4,786,170 A	11/1988	Groeblner	5,447,841 A	9/1995	Gray et al.
4,786,684 A	11/1988	Glass	5,474,796 A	12/1995	Brennan
4,794,150 A	12/1988	Steel	5,486,452 A	1/1996	Gordon et al.
4,808,508 A	2/1989	Platzer	5,489,507 A	2/1996	Chehab
4,810,869 A	3/1989	Yabe et al.	5,489,678 A	2/1996	Fodor et al.
4,811,062 A	3/1989	Tabata et al.	5,492,806 A	2/1996	Drmanac et al.
4,812,512 A	3/1989	Buendia et al.	5,494,810 A	2/1996	Barany et al.
4,820,630 A	4/1989	Taub	5,510,270 A	4/1996	Fodor et al.
4,822,566 A	4/1989	Newman	5,525,464 A	6/1996	Drmanac et al.
4,833,092 A	5/1989	Geysen	5,527,681 A	6/1996	Holmes
4,844,617 A	7/1989	Kelderman et al.	5,552,270 A	9/1996	Khrapko et al.
4,846,552 A	7/1989	Veldkamp et al.	5,556,961 A	9/1996	Foot et al. 536/27.1
4,849,513 A	7/1989	Smith et al.	5,561,071 A	10/1996	Hollenberg et al. 437/1
4,855,225 A	8/1989	Fung et al.	5,571,639 A	11/1996	Hubbell et al.
4,865,990 A	9/1989	Stead et al.	5,593,839 A	1/1997	Hubbell et al.
4,868,103 A	9/1989	Stavrianopoulos et al.	5,643,728 A	7/1997	Slater et al.
4,874,500 A	10/1989	Madou et al.	5,653,939 A	8/1997	Hollis et al.
4,886,741 A	12/1989	Schwartz	5,667,667 A	9/1997	Southern
4,888,278 A	12/1989	Singer et al.	5,667,972 A	9/1997	Drmanac et al.
4,923,901 A	5/1990	Koester et al.	5,695,940 A	12/1997	Drmanac et al.
4,925,785 A	5/1990	Wang et al.	5,698,393 A	12/1997	Macioszek et al.
4,946,942 A	8/1990	Fuller et al.	5,700,637 A	12/1997	Southern
4,965,188 A	10/1990	Mullis et al.	5,707,806 A	1/1998	Shuber
4,973,493 A	11/1990	Guire	5,744,305 A	4/1998	Fodor et al.
4,979,959 A	12/1990	Guire	5,776,737 A	7/1998	Dunn
4,981,783 A	1/1991	Augenlicht	5,777,888 A	7/1998	Rine et al.
4,981,985 A	1/1991	Kaplan et al.	5,800,992 A	9/1998	Fodor et al.
4,984,100 A	1/1991	Takayama et al.	5,807,552 A	9/1998	Brown et al.
4,987,065 A	1/1991	Stavrianopoulos et al.	5,830,645 A	11/1998	Pinkel et al.
4,988,617 A	1/1991	Landegren et al.	5,843,767 A	12/1998	Beattie
4,992,383 A	2/1991	Farnsworth	5,846,708 A	12/1998	Hollis et al.
4,994,373 A	2/1991	Stavrianopoulos et al.	5,871,697 A	2/1999	Rothberg et al.
5,002,867 A	3/1991	Macevicz	6,054,270 A	4/2000	Southern
5,021,550 A	6/1991	Zeiger			
5,026,773 A	6/1991	Steel			
5,026,840 A	6/1991	Dattagupta et al.			
5,028,525 A	7/1991	Gray et al.			
5,043,265 A	8/1991	Tanke et al.			
5,047,524 A	9/1991	Andrus et al.			
5,079,600 A	1/1992	Schnur et al.			
5,081,584 A	1/1992	Omichinski et al.			
5,082,830 A	1/1992	Brakel et al.			
5,091,652 A	2/1992	Mathies et al.			
5,112,962 A	5/1992	Letsinger et al.			
5,141,813 A	8/1992	Nelson			
5,143,854 A	9/1992	Pirung et al.			
5,149,625 A	9/1992	Church et al.			
5,153,319 A	10/1992	Caruthers et al.			
5,192,980 A	3/1993	Dixon et al.			
5,200,051 A	4/1993	Cozzette et al.			
5,202,231 A	4/1993	Drmanac et al.			
5,206,137 A	4/1993	Ip et al.			
5,215,882 A	6/1993	Bahl et al.			
5,215,889 A	6/1993	Schultz			
5,225,326 A	7/1993	Barany et al.			
5,232,829 A	8/1993	Longiaru et al.			
5,235,028 A	8/1993	Barany et al.			
5,242,974 A	9/1993	Holmes			
5,252,743 A	10/1993	Barrett et al.			
5,256,549 A	10/1993	Urdea et al.			
5,258,506 A	11/1993	Urdea et al.			
5,306,641 A	4/1994	Saccocio			
5,310,893 A	5/1994	Erlich et al.			
5,324,633 A	6/1994	Fodor et al.			
5,348,855 A	9/1994	Dattagupta et al.			
5,384,261 A	1/1995	Winkler et al.			
5,405,783 A	4/1995	Pirung et al.			
5,424,186 A	6/1995	Fodor et al.			
5,436,327 A	7/1995	Southern et al.			

FOREIGN PATENT DOCUMENTS

EP	046 083	2/1982
EP	088 636	9/1983
EP	103 197	3/1984
EP	127 438	12/1984
EP	063 810	3/1986
EP	194 132	9/1986
EP	228 075	7/1987
EP	245 662	11/1987
EP	268 237	5/1988
EP	281 927	9/1988
EP	288 310	10/1988
EP	304 202	2/1989
EP	307 476	3/1989
EP	319 012	6/1989
EP	328 256	8/1989
EP	333 561	9/1989
EP	228 310	10/1989
EP	337 498	10/1989
EP	386 229	4/1990
EP	373 203	6/1990
EP	292 546	10/1990
EP	173 339	1/1992
EP	171 150	3/1992
EP	237 362	3/1992
EP	185 547	6/1992
EP	260 634	6/1992
EP	232 967	4/1993
EP	235 726	5/1993
EP	476 014	8/1994
EP	225 807	10/1994
EP	717 113	6/1996
EP	721 016	7/1996
EP	848 067	6/1998
EP	619 321	1/1999
FR	2559783	3/1988

US 6,420,169 B1

Page 3

GB	2156074	3/1988
GB	2196476	4/1988
GB	8810400.5	5/1988
GB	2233654	1/1991
GB	2248840	9/1992
JP	49-110601	10/1974
JP	60-248669	12/1985
JP	63-084499	4/1988
JP	63-223557	9/1988
JP	1-233447	9/1989
NO	P 913186	8/1991
WO	WO 84/03151	8/1984
WO	WO 84/03564	9/1984
WO	WO 85/01051	3/1985
WO	WO 86/00991	2/1986
WO	WO 86/06487	11/1986
WO	WO 97/10977	5/1988
WO	WO 88/04777	6/1988
WO	WO 89/05616	6/1989
WO	WO 89/08834	9/1989
WO	WO 89/10977	11/1989
WO	WO 89/11548	11/1989
WO	WO 89/12819	12/1989
WO	WO 90/00626	1/1990
WO	WO 90/00887	2/1990
WO	WO 90/15070	2/1990
WO	WO 90/03382	4/1990
WO	WO 90/04652	5/1990
WO	WO 91/04266	4/1991
WO	WO 91/07087	5/1991
WO	WO 92/16655	1/1992
WO	WO 92/10092	6/1992
WO	WO 92/10588	6/1992
WO	WO 93/02992	2/1993
WO	WO 93/09668	5/1993
WO	WO 88/01302	6/1993
WO	WO 93/11262	6/1993
WO	WO 93/17126	9/1993
WO	WO 93/22456	11/1993
WO	WO 93/22480	11/1993
WO	WO 95/00530	1/1995
WO	WO 95/11995	5/1995
WO	WO 95/33846	12/1995
WO	WO 96/23078	8/1996
WO	WO 97/10365	3/1997
WO	WO 97/17317	5/1997
WO	WO 97/19410	5/1997
WO	WO 97/27317	7/1997
WO	WO 97/29212	8/1997
WO	WO 97/31256	8/1997
WO	WO 97/45559	12/1997
WO	WO 98/03676	1/1998
WO	WO 98/31836	7/1998

OTHER PUBLICATIONS

"A Sequencing Reality Check," *Science*, 242:1245 (1988).

"Affymax raises \$25 million to develop high-speed drug discovery system," *Biotechnology News*, 10(3):7-8.

"Preparation of fluorescent-labeled DNA and its use as a probe in molecular hybridization," *Bioorg Khim*, 12(11):1508-1513 (1986).

Abbott et al., "Manipulation of the Wettability of Surfaces on the 0.1-to-Micrometer Scale Through Micromachining and Molecular Self-Assembly," *Science*, 257:1380-1382 (1992).

Adams et al., "Complementary DNA Sequencing: Expressed Sequence Tags and Human Genome Project," *Science*, 252(5013):1651-1656 (1991).

Adams et al., "Photolabile Chelators That "Cage" Calcium with Improved Speed of Release and Pre-Photolysis Affinity," *J. Gen. Physiol.*, p. 9a (12/86).

Adams et al., "Biologically Useful Chelators That Take Up Ca²⁺ upon Illumination," *J. Am. Chem. Soc.*, 111:7957-7968 (1989).

Amit et al., "Photosensitive Protecting Groups of Amino Sugars and Their Use in Glycoside Synthesis. 2-Nitrobenzyloxycarbonylamino and 6-Nitroveratryloxycarbonylamino Derivatives," *J. Org. Chem.*, 39(2):192-196 (1974).

Amit et al., "Photosensitive Protecting Groups—A Review," *Israel J. Chem.*, 12(1-2):103-113 (1974).

Applied Biosystems, Model 431A Peptide Synthesizer User's manual, Sections 2 and 6, (Aug. 15, 1989).

Ajayaghosh et al., "Solid-Phase Synthesis of N-Methyl- and N-Ethylamides of Peptides Using Photolytically Detachable

((3-Nitro-4((alkylamino)methyl)benzamido)methyl)polystyrene Resin," *J. Org. Chem.*, 55(9):2826-2829 (1990).

Ajayaghost et al., "Solid-phase synthesis of C-terminal peptide amides using a photoremovable α -methylphenacylamido anchoring linkage," *Proc. Ind. Natl. Sci (Chem.Sci.)*, 100(5):389-396 (1988).

Ajayaghosh et al., "Polymer-supported Solid-phase Synthesis of C-Terminal Peptide N-Methylamides Using a Modified Photoremovable 3-Nitro-4-N-methylaminomethylpolystyrene Support," *Ind. J. Chem.*, 27B:1004-1008 (1988).

Ajayaghosh et al., "Polymer-Supported Synthesis of Protected Peptide Segments on a Photosensitive o-Nitro(α -Methyl)Bromobenzyl Resin," *Tetrahedron*, 44(21):6661-6666 (1988).

Arnold et al., "A Novel Universal Support for DNA & RNA Synthesis," abstract from *Federation Proceedings*, 43(7):abstract No. 3669 (1984).

Atherton et al., *Solid Phase Peptide Synthesis: A Practical Approach*, IRL Press, (1989), tbl. of cont., pp. vii-ix.

Augenlicht et al., "Cloning and Screening of Sequences Expressed in a Mouse Colon Tumor," *Cancer Research*, 42:1088-1093 (1982).

Augenlicht et al., "Expression of Cloned Sequences in Biopsies of Human Colonic Tissue and in Colonic Carcinoma Cells Induced to Differentiate in Vitro," *Cancer Res.*, 47:6017-6021 (1987).

Bains, W., "Hybridization Methods for DNA Sequencing," *Genomics*, 11(2):294-301 (1991).

Bains et al., "A Novel Method for Nucleic Acid Sequences Determination," *J. Theor. Biol.*, 135:303-307 (1988).

Bains, W., "Alternative Routes Through the Genome," *Biotechnology*, 8:1251-1256 (1988).

Balachander et al., "Functionalized Siloxy-Anchored Monolayers with Exposed Amino, Azido, Bromo, or Cyano Groups," *Tetrahed. Ltrs.*, 29(44):5593-5594 (1988).

Baldwin et al., "New Photolabile Phosphate Protecting Groups," *Tetrahed.*, 46(19):6879-6884 (1990).

Barltrop et al., "Photosensitive Protective Groups," *Chemical Communications*, pp. 822-823 (1966).

Barinaga, M., "Will 'Will DNA Chip' Speed Genome Initiative," *Science*, 253:1489 (1985).

Bart et al., "Microfabricated Electrohydrodynamic Pumps," *Sensors and Actuators*, A21-A23:193-197 (1990).

US 6,420,169 B1

Page 4

- Bartsh et al., "Cloning and mRNA sequences from the human colon: Preliminary characterisation of defined mRNAs in normal and neoplastic tissues," *Br.J.Can.*, 54:791-798 (1986).
- Baum, R., "Fledgling firm targets drug discovery process," *Chem. Eng. News*, p. 10-11 (1990).
- Beltz et al., "Isolation of Multigene Families and Determination of Homologies by Filter Hybridization Methods," *Methods in Enzymology*, 100:266-285 (1983).
- Benschop, Chem.ABSTRACTS 114(26):256643 (1991).
- Bhatia et al., "New Approach To Producing Patterned Biomolecular Assemblies," *J. American Chemical Society*, 114:4432-4433 (1992).
- Biorad Chromatography Electrophoresis Immunochemistry Molecular Biology HPLC catalog M 1987 pp. 182.
- Blawas et al., "Step-and-Repeat Photopatterning of Protein Features Using Caged-Biotin-BSA: Characterization and Resolution," *Langmuir*, 14(15):4243-4250 (1998).
- Blawas, A.S., "Photopatterning of Protein Features using Caged-biotin-Bovine Serum Albumin," dissertation for Ph.D at Duke University in 1998.
- Bos et al., "Amino-acid substitutions at codon 13 of the N-ras oncogene in human acute myeloid leukaemia," *Nature*, 315:726-730 (1985).
- Boyle et al., "Differential distribution of long and short interspersed element sequences in the mouse genome: Chromosome karyotyping by fluorescence in situ hybridization," *PNAS*, 87:7757-7761 (1990).
- Brock et al., "Rapid fluorescence detection of in situ hybridization with biotinylated bovine herpesvirus-1 DNA probes," *J.Veterinary Diagnostic Invest.*, 1:34-38 (1989).
- Burgi et al., "Optimization in Sample Stacking for High-Performance Capillary Electrophoresis," *Anal. Chem.*, 63:2042-2047 (1991).
- Cameron et al., "Photogeneration of Organic Bases from o-Nitrobenzyl-Derived Carbamates," *J. Am. Chem. Soc.*, 113:4303-4313 (1991).
- Carrano et al., "A High-Resolution, Fluorescence-Based, Semiautomated Method for DNA Fingerprinting," *Genomics*, 4:129-136 (1989).
- Caruthers, M.H., "Gene Synthesis Machines: DNA Chemistry and Its Uses," *Science*, 230:281-285 (1985).
- Chatterjee et al., "Inducible Alkylation of DNA Using an Oligonucleotide-Quinone Conjugate," *Am. J. Chem. Soc.*, 112:6397-6399 (1990).
- Chee et al., "Accessing Genetic Information with High-Density DNA Arrays," *Science*, 274:610-614 (1996).
- Chehab et al., "Detection of sickle cell anaemia mutation by colour DNA amplification," *Lancet*, 335:15-17 (1990).
- Chehab et al., "Detection of specific DNA sequences by fluorescence amplification: A color complementation assay," *PNAS*, 86:9178-9182 (1989).
- Clevite Corp., Piezoelectric Technology, Data for Engineers.
- Corbett et al., "Reaction of Nitroso Aromatics with Glyoxylic Acid. A New Path to Hydroxamic Acids," *J. Org. Chem.*, 45:2834-2839 (1980).
- Craig et al., "Ordering of cosmid clones covering the Herpes simplex virus type 1 (HSV-1) genome: a test case for fingerprinting by hybridization," *Nuc. Acid. Res.*, 18(9):2653-2660 (1990).
- Cummings et al., "Photoactivable Fluorophores. 1. Synthesis and Photoactivation of o-Nitrobenzyl-Quenched Fluorescent Carbamates," *Tetrahedron Letters*, 29(1):65-68 (1988).
- Diggelmann, "Investigating the VLSIPS synthesis process," Sep. 9, 1994.
- Di Mauro et al., "DNA Technology in Chip Construction," *Adv. Mater.*, 5(5):384-386 (1993).
- Drmanac et al., "Partial Sequencing by Oligo-Hybridization Concept and Application in Genome Analysis," 1st Int. Conf. Electrophor., Supercomp., Hum. Genome pp. 60-74 (1990).
- Drmanac et al., "Sequencing by Oligonucleotide Hybridization: A Promising Framework in Decoding of the Genome Program?," 1st Int. Conf. Electrophor., Supercomp., Hum. Genome pp. 47-59 (1990).
- Drmanac et al., "Laboratory Methods, Reliable Hybridization of Oligonucleotides as Short as Six Nucleotides," *DNA and Cell Biol.*, 9(7):527-534 (1990).
- Drmanac et al., "Sequencing of Megabase Plus DNA by Hybridization: theory of the Method," *Genomics*, 4:114-128 (1989).
- Dramanac et al., "Sequencing of Megabase Plus DNA by Hybridization: Theory of the Method," abstract of presentation given at Cold Spring Harbor Symposium on Genome Mapping and Sequencing, Apr. 27, 1988 thru May 1, 1988.
- Dulcey et al., "Deep UV Photochemistry of Chemisorbed Monolayers: Patterned Coplanar Molecular Assemblies," *Science*, 252:551-554 (1991).
- Duncan et al., "Affinity Chromatography of a Sequence-Specific DNA Binding Protein Using Teflon-Linked Oligonucleotides," *Analytical Biochemistry*, 169:104-108 (1988).
- Effenhauser et al., "Glass Chips for High-speed Capillary Electrophoresis Separations with Submicrometer Plate Heights," *Anal. Chem.*, 65:2637-2642 (1993).
- Effenhauser et al., "High-Speed Separation of Antisense Oligonucleotides on a Micromachined Capillary Electrophoresis Device," *Anal. Chem.*, 66:2949-2953 (1994).
- Ekins et al., "High Specific Activity Chemiluminescent and Fluorescent Markers: their Potential Application to High Sensitivity and 'Multi-analyte' Immunoassays," *J. Bioluminescence Chemiluminescence*, 4:59-78 (1989).
- Ekins et al., "Development of Microspot Multi-Analyte Radiometric Immunoassay Using dual Fluorescent-Labelled Antibodies," *Anal. Chimica Acta*, 227:73-96 (1989).
- Ekins et al., "Multianalyte Microspot Immunoassay-Microanalytical 'Compact Disk' of the Future," *Clin. Chem.*, 37(11):1955-1967 (1991).
- Ekins, R.P., "Multi-Analyte immunoassay*," *J. Pharmaceutical Biomedical Analysis*, 7(2):155-168 (1989).
- Ekins et al., "Fluorescence Spectroscopy and its Application to a New Generation of High Sensitivity, Multi-Microspot, Multianalyte, Immunoassay," *Clin. Chim. Acta*, 194:91-114 (1990).
- Evans et al., "Microfabrication for Automation of Molecular processes in Human Genome Analysis," *Clin. Chem.*, 41(11):1681 (1995).
- Evans et al., "Physical mapping of complex genomes by cosmid multiplex analysis," *PNAS*, 86:5030-5034 (1989).
- Ezaki et al., "Small-Scale DNA Preparation for Rapid Genetic Identification of *Campylobacter* Species without Radioisotope," *Microbiol. Immunology*, 32(2):141-150 (1988).
- Fan et al., "Mapping small DNA sequences by fluorescence in situ hybridization directly on banded metaphase chromosomes," *PNAS*, 87(16):6223-6227 (1990).

US 6,420,169 B1

Page 5

- Fan et al., "Micromachining of Capillary Electrophoresis Injectors and Separators on Glass Chips and Evaluation of Flow at Capillary Intersections," *Anal. Chem.*, 66:177-184 (1994).
- Fettinger et al., "Stacked modules for micro flow systems in chemical analysis: concept and studies using an enlarged model," *Sensors and Actuators*, B17:19-25 (1993).
- Flanders et al., "A new interferometric alignment technique," *App. Phys. Ltrs.*, 31(7):426-429 (1977).
- Fodor et al., "Multiplexed biochemical assays with biological chips," *Nature*, 364:555-556 (1993).
- Fodor et al., "Light-directed, Spatially Addressable Parallel Chemical Synthesis," *Science*, 251:767-773 (1991).
- Forman et al., "Thermodynamics of Duplex Formation and Mismatch Discrimination on Photolithographically Synthesized Oligonucleotide Arrays," chapter 13pp. 206-228 from *Molecular Modeling of Nucleic Acids*, ACS Symposium Series 682, 4/13-17/97, Leontis et al., eds.
- Frank et al., "Simultaneous Multiple Peptide Synthesis Under Continuous flow Conditions on Cellulose Paper Discs as Segmental Solid Supports," *Tetrahedron*, 44(19):6031-6040 (1988).
- Frank et al., "Automation of DNA Sequencing Reactions and Related Techniques: A Workstation for Micromanipulation of Liquids," *Bio/Technology*, 6:1211-1212 (1988).
- Frank et al., "Simultaneous Synthesis and Biological Application of DNA Fragments: An Efficient and Complete Methodology," *Methods in Enzymology*, 154:221-250 (1987).
- Fuhr et al., "Travelling wave-driven microfabricated electrohydrodynamic pumps for liquids," *J. Micromech. Microeng.*, 4:217-226 (1994).
- Fuller et al., "Urethane-Protected Amino Acid N-Carboxy Anhydrides and Their Use in Peptide Synthesis," *J. Amer. Chem. Soc.*, 112(20):7414-7416 (1990).
- Furka et al., "General method for rapid synthesis of multi-component peptide mixtures," *Int. J. Peptide Protein Res.*, 37:487-493 (1991).
- Furka et al., "Cornucopia of Peptides by Synthesis," 14th Int. Congress of Biochem. abst.# FR:013, 7/10-15/88 Prague, Czechoslovakia.
- Furka et al., "More Peptides by Less Labour," abst. 288, Int. Symp. Med. Chem., Budapest Hungary 8/15-19/88.
- Gait, eds., pp. 1-115 from *Oligonucleotide Synthesis: A Practical Approach*, IRL Press, (1984).
- Gazard et al., "Lithographic Technique Using Radiation-Induced Grafting of Acrylic Acid into Poly(Methyl Methacrylate) Films," *Polymer Engineering and Science*, 20(16):1069-1072 (1980).
- Gergen et al., "Filter replicas and permanent collections of recombinant DNA plasmids," *Nuc.Acids Res.*, 7(8):2115-2137 (1979).
- Getzoff et al., "Mechanisms of Antibody Binding to a Protein," *Science*, 235:1191-1196 (1987).
- Geysen et al., "Strategies for epitope analysis using peptide synthesis," *J. Immunol. Meth.*, 102:259-274 (1987).
- Geysen et al., "Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid," *PNAS*, 81:3998-4002 (1984).
- Geysen et al., "A synthetic strategy for epitope mapping," from *Peptides:Chem. & Biol.*, Proc. of 10th Am. Peptide Symp., 5/23-28/87, pp. 519-523, (1987).
- Geysen, "Antigen-antibody interactions at the molecular level: adventures in peptide synthesis," *Immunol. Today*, 6(12):364-369 (1985).
- Geysen et al., "Cognitive Features of Continuous Antigenic Determinants," from *Synthetic Peptides: Approaches to Biological Probes*, p. 19-30, (1989).
- Geysen et al., "Chemistry of Antibody Binding to a Protein," *Science*, 235:1184-1190 (1987).
- Geysen et al., "The delineation of peptides able to mimic assembled epitopes," 1986 CIBA Symp., pp. 130-149.
- Geysen et al., "Cognitive Features of Continuous Antigenic Determinants," *Mol. Recognit.*, 1(1):1-10 (1988).
- Geysen et al., "A Prio Ri Delineation of a Peptide Which Mimics A Discontinuous Antigenic Determinant," *Mol. Immunol.*, 23(7):709-715 (1986).
- Gilon et al., "Backbone Cyclization: A New Method for Conferring Conformational Constraint on Peptides," *Biopolymers*, 31(6):745-750 (1991).
- Gingeras et al., "Hybridization properties of immobilized nucleic acids," *Nuc. Acids Res.*, 15(13):5373-5390.
- Gummerlock et al., "RAS Enzyme-Linked Immunoblot Assay Discriminates p21 Species: A Technique to Dissect Gene Family Expression," *Anal. Biochem.*, 180:158-168 (1989).
- Gurney et al., "Activation of a potassium current by rapid photochemically generated step increases of intracellular calcium in rat sympathetic neurons," *PNAS*, 84:3496-3500 (1987).
- Haase et al., "Detection of Two Viral Genomes in Single Cells by Double-Label Hybridization in Situ and Color Microradioautography," *Science*, 227:189-192 (1985).
- Hacia, et al., "Two color hybridization analysis using high density oligonucleotide arrays and energy transfer dyes," *Nuc. Acids Res.*, 26(16):3865-3866 (1998).
- Hack, M.L., "Conics Formed to Make Fluid & Industrial Gas Micromachines," *Genetic Engineering News*, 15(18):1, 29 (1995).
- Hagedorn et al., "Pumping of Water Solutions in Microfabricated Electrohydrodynamic Systems," from *Micro Electro Mechanical Systems conference in Travemunde Germany* (1992).
- Hames et al., *Nuclear acid hybridization, a practical approach*, cover page and table of contents (1985).
- Hanahan et al., "Plasmid Screening at High Colony Density," *Meth. Enzymology*, 100:333-342 (1983).
- Hanahan et al., "Plasmid screening at high colony density," *Gene*, 10:63-67 (1980).
- Haridasan et al., "Peptide Synthesis using Photolytically Cleavable 2-Nitrobenzyloxycarbonyl Protecting Group," *Proc. Indian Natn. Sci. Adad.*, 53A(6):717-728 (1987).
- Harrison et al., "Capillary Electrophoresis and Sample Injection Systems Integrated on a Planar Glass Chip," *Anal. Chem.*, 64:1926-1932 (1992).
- Harrison et al., "Micromachining a Minaturized Capillary Electrophoresis-Based Chemical Analysis System on a Chip," *Science*, 261:895-897 (1993).
- Harrison et al., "Towards minaturized electrophoresis and chemical analysis systems on silicon: an alternative to chemical sensors*," *Sensors and Actuators*, B10:107-116 (1993).
- Harrison et al., "Rapid separation of fluorescein derivatives using a micromachined capillary electrophoresis system," *Analytica Chimica Acta*, 283:361-366 (1993).
- Hellberg et al., "Minimum analogue peptide sets (MAPS) for quantitative structure-activity relationships," *Int. J. Peptide Protein Res.*, 37:414-424 (1991).

US 6,420,169 B1

Page 6

- Hilser et al., "Protein and peptide mobility in capillary zone electrophoresis, A comparison of existing models and further analysis," *J. Chromatography*, 630:329-336 (1993).
- Ho et al., "Highly Stable Biosensor Using an Artificial Enzyme," *Anal. Chem.*, 59:536-537 (1987).
- Hochgeschwender et al., "Preferential expression of a defined T-cell receptor β -chain gene in hapten-specific cytotoxic T-cell clones," *Nature*, 322:376-378 (1986).
- Hodgson, J., "Assays A La Photolithography," *Biotech.*, 9:419 (1991).
- Hopman et al., "Bi-color detection of two target DNAs by non-radioactive in situ hybridization*," *Histochem.*, 85:1-4 (1986).
- Iwamura et al., "1-Pyrenylmethyl Esters, Photolabile Protecting Groups for Carboxylic Acids," *Tetrahedron Ltrs.*, 28(6):679-682 (1987).
- Iwamura et al., "1-(α -Diazobenzyl)pyrene: A Reagent for Photolabile and Fluorescent Protection of Carboxyl Groups of Amino Acids and Peptides," *Synlett*, p. 35-36 (1991).
- Jacobson et al., "Effects of Injection Schemes and Column Geometry on the Performance of Microchip Electrophoresis Devices," *Anal. Chem.*, 66:1107-1113 (1994).
- Jacobsen et al., "Open Channel Electrochromatography on a Microchip," *Anal. chem.*, 66:2369-2373 (1994).
- Jacobson et al., "Microchip Capillary Electrophoresis with an Integrated Postcolumn Reactor" *Anal. Chem.*, 66:3472-3476 (1994).
- Jacobson et al., "Precolumn Reactions with Electrophoretic Analysis Integrated on a Microchip," *Anal. Chem.*, 66:4127-4132 (1994).
- Jacobson et al., "Microfabricated chemical measurement systems," *Nature Medicine*, 1(10):1093-1096 (1995).
- Jacobsen et al., "Fused Quartz Substrates for Microchip Electrophoresis," *Anal. chem.*, 67:2059-2063 (1995).
- Jacobson et al., "High-Speed Separations on a Microchip," *Anal. Chem.*, 66:1114-1118 (1994).
- Jacobson et al., "Microchip electrophoresis with sample stacking," *Electrophoresis*, 16:481-486 (1995).
- Jayakumari, "Peptide synthesis in a triphasic medium catalysed by papain immobilized on a crosslinked polystyrene support," *Indian J. Chemistry*, 29B:514-517 (1990).
- Kaiser et al., "Peptide and Protein Synthesis by Segment Synthesis-Condensation," *Science*, 243:187-192 (1989).
- Kaplan et al., "Photolabile chelators for the rapid photorelease of divalent cations," *PNAS*, 85:6571-6575 (1988).
- Karube, "Micro-biosensors based on silicon fabrication technology," chapter 25 from *Biosensors: Fundamentals and Applications*, Turner et al., eds., Oxford Publ., 1987, pp. 471-480 (1987).
- Kates et al., "A Novel, Convenient, Three-dimensional Orthogonal Strategy for Solid-Phase Synthesis of Cyclic Peptides 1-3," *Tetrahed. Letters*, 34(10):1549-1552 (1993).
- Kerkof et al., "A Procedure for Making Simultaneous Determinations of the Relative Levels of Gene Transcripts in Tissue or Cells," *Anal. Biochem.*, 188:349-355 (1990).
- Khrapko et al., "An Oligonucleotide hybridization approach to DNA sequencing," *FEBS Lett.*, 256(1,2):118-122 (1989).
- Kievits et al., "Rapid subchromosomal localization of cosmids by nonradioactive in situ hybridization," *Cytogenetics Cell Genetics*, 53(2-3):134-136 (1990).
- Kimura et al., "Immobilized Enzyme Membrane Fabrication Method using an Ink Jet Nozzle," *Biosensors*, 4:41-52 (1988).
- Kimura et al., "An Integrated SOS/FET Multi-Biosensor," *Sensors & Actuators*, 9:373-387 (1986).
- Kitazawa et al., "In situ DNA-RNA hybridization using in vivo bromodeoxyuridine-labeled DNA probe," *Histochemistry*, 92:195-199 (1989).
- Kleinfeld et al., "Controlled Outgrowth of Dissociated Neurons on Patterned Substrates," *J. Neurosci.*, 8(11):4098-4120 (1988).
- Knight, P., "Materials and Methods/Microsequencers for Proteins and Oligosaccharides," *Bio/Tech.*, 7:1075-76 (1989).
- Kohara et al., "The Physical Map of the Whole *E. coli* Chromosome: Application of a New Strategy for Rapid Analysis and Sorting of a Large Genomic Library," *Cell*, 50:495-508 (1987).
- Krile et al., "Multiplex holography with chirp-modulated binary phase-coded reference-beam masks," *Applied Opt.*, 18(1):52-56 (1979).
- Labat, I., "Subfragments as an informative characteristic of the DNA molecule—computer simulation," research report submitted to the University of Belgrade College of Natural Sciences and Mathematics, (1988).
- Lainer et al., "Human Lymphocyte Subpopulations Identified by Using Three-Color Immunofluorescence and Flow Cytometry Analysis: Correlation of Leu-2, Leu-3, Leu-7, Leu-8, and Leu-11 Clee Surface Antigen Expression," *Journal of Immunology*, 132(1):151-156 (1984).
- Lam et al., "A new type of synthetic peptide library for identifying ligand-binding activity," *Nature*, 354:82-84 (1991).
- Laskey et al., "Messenger RNA prevalence in sea urchin embryos measured with cloned cDNAs," *PNAS*, 77(9):5317-5321 (1980).
- Lee et al., "synthesis of a Polymer Surface Containing Covalently Attached Triethoxysilane Functionality: Adhesion to Glass," *Macromolecules*, 21:3353-3356 (1988).
- Lehrach et al., "Labelling oligonucleotides to high specific activity (I)," *Nuc. Acids Res.*, 17(12):4605-4610 (89).
- Lehrach et al., "Phage Vectors—EMBL Series," *Meth. Enzymology*, 153:103-115 (1987).
- Levy, M.F., "Preparing Additive Printed Circuits," *IBM Tech. Discl. Bull.*, 9(11):1473 (1967).
- Lichter et al., "High-Resolution Mapping of Human Chromosome 11 by in Situ hybridization with Cosmid Clones," *Science*, 247:64-69 (1990).
- Lichter et al., "Fluorescence in situ hybridization with Alu and L1 polymerase chain reaction probes for rapid characterization of human chromosomes in hybrid cell lines," *PNAS*, 87:6634-6638 (1990).
- Lichter et al., "Rapid detection of human chromosome 21 aberrations by in situ hybridization," *PNAS*, 85:9664-9668 (1988).
- Lichter et al., "Is non-isotopic in situ hybridization finally coming of age," *Nature*, 345:93-94 (1990).
- Lieberman et al., "A Light source Smaller Than the Optical Wavelength," *Science*, 247:59-61 (1990).
- Lipshutz et al., "Using Oligonucleotide Probe Arrays To Access Genetic Diversity," *BioTech.*, 19(3):442-7 (1995).
- Liu et al., "Sequential Injection Analysis in Capillary Format with an Electroosmotic Pump," *Talanta*, 41(11):1903-1910 (1994).
- Lockhart et al., "Expression monitoring by hybridization to high-density oligonucleotide arrays," *Nat. Biotech.*, 14:1675-1680 (1996).

US 6,420,169 B1

Page 7

- Logue et al., "General Approaches to Mask Design for Binary Optics," *SPIE*, 1052:19–24 (1989).
- Loken et al., "three-color Immunofluorescence Analysis of Leu Antigens on Human Peripheral Blood Using Two Lasers on a Fluorescence-Activated Cell Sorter," *Cytoetry*, 5:151–158 (1984).
- Love et al., "Screening of λ Library for Differentially Expressed Genes Using in Vitro Transcripts," *Anal. Biochem.*, 150:429–441 (1985).
- Lowe, C.R., "Biosensors," *Trends in Biotech.*, 2:59–65 (1984).
- Lowe, C.R., "An Introduction to the Concepts and Technology of Biosensors," *Biosensors*, 1:3–16 (1985).
- Lowe, C. R., *Biotechnology and Crop Improvement and Protection*, BCPC Publications, pp. 131–138 (1986).
- Lowe et al., "Solid-Phase Optoelectronic Biosensors," *Methods in Enzymology*, 137:338–347 (1988).
- Lowe, C.R., "Biosensors," *Phil. Tran. R. Soc. Lond.*, 324:487–496 (1989).
- Lu et al., "Differential screening of murine ascites cDNA libraries by means of in vitro transcripts of cell-cycle-phase-specific cDNA and digital image processing," *Gene*, 86:185–192 (1990).
- Lysov et al., "A new method for determining the DNA nucleotide sequence by hybridization with oligonucleotides," *Doklady Biochem.*, 303(1–6):436–438 (1989).
- Lysov et al., "DNA Sequencing by Oligonucleotide Hybridization," First International Conference on Electrophoresis, Supercomputing and the Human Genome, 4/10–13/90 p. 157.
- MacDonald et al., "A Rapid ELISA for Measuring Insulin in a Large Number of Research Samples," *Metabolism*, 38(5):450–452 (1989).
- Mairanovsky, V.G., "Electro-Deprotectin-Electrochemical Removal of Protecting Groups**," *Agnew. Chem. Int. Ed. Engl.*, 15(5):281–292 (1976).
- Manz et al., "Miniaturized Total Chemical Analysis Systems: a Novel Concept for Chemical Sensing," *Sensors and Actuators*, B1:244–248 (1990).
- Manz et al., "Micromachining of monocrystalline silicon and glass for chemical analysis systems, A look into next century's technology or just a fashionable craze?," *Trends in Analytical Chem.*, 10(5):144–149 (1991).
- Manz et al., "Planar chips technology for miniaturization and integration of separation techniques into monitoring systems, Capillary electrophoresis on a chip," *J. Chromatography*, 593:253–258 (1992).
- Manz et al., "Planar Chips Technology for Miniaturization of Separation Systems: A Developing Perspective in Chemical Monitoring," chapter 1, 1–64 (1993).
- Manz et al., "Electroosmotic pumping and electrophoretic separations for minaturized chemical analysis systems," *J. Micromech. Microeng.*, 4:257–265 (1994).
- Masiakowski et al., "Cloning of cDNA sequences of hormone-regulated genes from the MCF-7 human breast cancer cell line," *Nuc. Acids Res.*, 10(24):7895–7903 (1982).
- Matsumoto et al., "Preliminary Investigation of Micropumping Based on Electrical Control of Interfacial Tension," *IEEE*, pp. 105–110 (1990).
- Matsuzawa et al., "Containment and growth of neuroblastoma cells on chemically patterned substrates," *J. Neurosci. Meth.*, 50:253–260 (1993).
- McCray et al., "Properties and Uses of Photoreactive Caged Compounds," *Ann. Rev. Biophys. Biophys. Chem.*, 18:239–270 (1989).
- McGall et al., "The Efficiency of Light-Directed Synthesis of DNA Arrays on Glass Substrates," *J. American Chem. Soc.*, 119(22):5081–5090 (1997).
- McGillis, VLSI Technology, Sze, eds., Chapter 7, "Lithography," pp. 267–301 (1983).
- McMurray, J.S., "Solid Phase Synthesis of a Cyclic Peptide Using Fmoc Chemistry," *Tetrahedron Letters*, 32(52):7679–7682 (1991).
- Mcinkoth et al., "Review: Hybridization of Nucleic Acids Immobilized on solid Supports," *Analytical Biochem.*, 138:267–284 (1984).
- Melcher et al., "Traveling-Wave Bulk Electroconvection Induced across a Temperature Gradient," *Physics of Fluids*, 10(6):1178–1185 (1967).
- Merrifield, R.B., "Solid Phase peptide Synthesis, I. The Synthesis of a Tetrapeptide," *J.Am.Chem.Soc.*, 85:2149–2154 (1963).
- Michiels et al., "Molecular approaches to genome analysis: a strategy for the construction of ordered overlapping clone libraries," *CABIOS*, 3(3):203–10 (1987).
- Mirzabekov, A.D., "DNA sequencing by hybridization—a megasequencing method and a diagnostic tool?," *TIBTECH*, 12:27–32 (1994).
- Monaco et al., "Human Genome Linking with Cosmids and Yeast Artificial Chromosomes", abstract from CSHS, p. 50, (1989).
- Morita et al., "Direct pattern fabrication on silicone resin by vapor phase electron beam polymerization," *J.Vac.Sci.Tech-nol.*, B1(4):1171–1173 (1983).
- Morrison et al., "Solution-Phase Detection of Polynucleotides Using Interacting Fluorescent Labels and Competitive Hybridization," *Anal. Biochem.*, 183:231–244 (1989).
- Munegumi et al., "thermal Synthesis of Polypeptides from N-Boc-Amino Acid (Aspartic Acid, β -Aminoglutaric Acid) Anhydrides," *Chem. Letters*, pp. 1643–1646 (1988).
- Mutter et al., "Impact of Conformation on the Synthetic Strategies for Peptide Sequences," pp. 217–228 from Chemistry of Peptides and Proteins, vol. 1, Proceedings of the Third USSR-FRG Symp., in USSR (1982).
- Nakamori et al., "A Simple and Useful Method for Simultaneous Screening of Elevated Levels of Expression of a Variety of Oncogenes in Malignant Cells," *Jpn. J. Cancer Res.*, 79:1311–1317 (1988).
- Nederlof et al., "Multiple Fluorescence In Situ Hybridization," *Cytometry*, 11:126–131 (1990).
- Nyborg, W., "Acoustic Streaming," chapter 11 pp. 265–329 from Physical Acoustics, Principles and Methods, Mason, eds., vol. II, part B, Academic Press, New York and London (1965).
- Oevirk et al., "High Performance Liquid Chromatography Partially Integrated onto a Silicon Chip," *Analyt. Meth. Instrumentation*, 2(2):74–82 (1995).
- Ohtsuka et al., "Studies on transfer ribonucleic acids and related compounds. IX Ribonucleic oligonucleotide synthesis using a photosensitive 0-nitrobenzyl protection at the 2'-hydroxyl group," *Nuc.Acids.Res.*, 1(10):1351–1357 (1974).
- Olefirowicz et al., "Capillary Electrophoresis for Sampling Single Nerve Cells," *Chimia*, 45(4):106–108 (1991).
- Patchornik et al., "Photosensitive Protecting Groups," *J.Am.Chem.Soc.*, 92(21):6333–6335 (1970).

US 6,420,169 B1

Page 8

- Patent Abstracts of Japan from EPO, Abst. 13:557, JP 1-233 447 (1989).
- Pease et al., "Light-generated oligonucleotide arrays for rapid DNA sequence analysis," *PNAS*, 91:5022-26 (1994).
- Pevzner, P.A., "1-Tuple DNA Sequencing: Computer Analysis," *J. Biomol. Struct. Dynam.*, 7(1):63-69 (1989).
- Pfahler et al., "Liquid Transport in Micron and Submicron Channels," *Sensors and Actuators*, A21-A23:431-4 (90).
- Pidgen et al., "Immobilized Artificial Membran Chromatography: Supports Composed of Membrane Lipids," *Anal. Biochem.*, 176:36-47 (89).
- Pillai, V.N. "Photoremovable Protecting Groups in Organic Synthesis," *Synthesis*, pp. 1-26 (1980).
- Pillai, V.N., "Photoremovable Protecting Groups in Organic Synthesis," *Synthesis*, pp. 1-26 (1980).
- Pillai et al., "3-Nitro-4-Aminomethylbenzoyl-derivate von Polyethylenglykolen: Eine neue Klasse von Photosensitiven loslichen Polymeren Tragern zur Synthese von C-terminalen Peptidamiden," *Tetrah. Lett.*, #36 p. 3409-3412 (1979).
- Pillai et al., "Synthetic Hydrophilic Polymers, Biomedical and Chemical Applications," *Naturwissenschaften*, 68:558-566 (1981).
- Pirrung et al., "Proofing of Photolithographic DNA Synthesis with 3',5'-Dimethoxybenzoinyloxycarbonyl-Protected Deoxynucleoside Phosphoramidites," *J. Org. Chem.*, 63(2):241-246 (1998).
- Pirrung et al., "Comparison of Methods for Photochemical Phosphoramidite-Based DNA Synthesis," *J. Org. Chem.*, 60:6270-6276 (1995).
- Ploax et al., "Cyclization of peptides on a solid support," *Int. J. Peptide Protein Research*, 29:162-169 (1987).
- Polsky-Cynkin et al., "Use of DNA Immobilized on Plastic and Agarose Supports to Detect DNA by Sandwich Hybridization," *Clin. Chem.*, 31(9):1428-1443 (1985).
- Poustka et al., "Molecular Approaches to Mammalian Genetics," *Cold Spring Harbor Symposia on Quantitative Biology*, 51:131-139 (1986).
- Purushothaman et al., "Synthesis of 4,5-diarylimidazole-2-thiones and their photoconversion to bis(4,5-diarylimidazol-2-yl) sulphides," *Ind. J. Chem.*, 29B:18-21 (1990).
- Quesada et al., "High-Sensitivity DNA Detection with a Laser-Excited Confocal Fluorescence Gel Scanner," *Biotechniques*, 10:616 (1991).
- Reichmanis et al., *J. Polymer Sci. Polymer Chem. Edition*, 23:1-8 (1985).
- Richter et al., "Electrohydrodynamic Micropump," *IEEE*, pp. 99-104 (1990).
- Richter et al., "Electrohydrodynamic Pumping and Flow Measurement," *IEEE*, pp. 271-276 (1991).
- Richter et al., "A Micromachined electrohydrodynamic (EHD) pump," *Sensors and Actuators*, A29:159-168 (91).
- Robertson et al., "A General and Efficient Route for Chemical Aminoacylation of Transfer RNAs," *J. Am. Chem. Soc.*, 113:2722-2729 (1991).
- Rodda et al., "The Antibody Response to Myoglobin-I. Systematic Synthesis of Myoglobin Peptides Reveals Location and Substructure of Species-Dependent Continuous Antigenic Determinants," *Mol. Immunol.*, 23(6):603-610 (1986).
- Rodgers, R.P., "Data Processing of Immunoassay Results," *Manual of Clin. Lab. Immunol.*, 3rd ed., ch. 15, pp. 82-87 (1986).
- Rose, D.J., "Free-solution reactor for post-column fluorescence detection in capillary zone electrophoresis," *J. Chromatography*, 540:343-353 (1991).
- Rovero et al., "Synthesis of Cyclic Peptides on solid Support," *Tetrahed. Letters*, 32(23):2639-2642 (1991).
- Sambrook, *Molecular Cloning—A Laboratory Manual*, publ. in 1989 (not included).
- Saiki et al., "Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes," *PNAS*, 86:6230-6234 (1989).
- Saiki et al., "Analysis of enzymatically amplified β -globin and HLA-DQ α DNA with Allele-specific oligonucleotide probes," *Nature*, 324:163-166 (1986).
- Scharf et al., "HLA class II allelic variation and susceptibility to pemphigus vulgaris," *PNAS*, 85(10):3504-3508 (1988).
- Urdea et al., "A comparison of non-radioisotopic hybridization assay methods using fluorescent, chemiluminescent and enzyme labeled synthetic oligodeoxyribonucleotide probes," *Nuc. Acids Res.*, 16(11):4937-4956 (1988).
- Van der Voort et al., "Design and Use of a Computer Controlled Confocal Microscope for Biological Applications," *Scanning*, 7(2):66-78 (1985).
- Van Hijfte et al., "Intramolecular 1,3-Diyl Trapping Reactions. A Formal Total Synthesis of -Corioline," *J. Organic Chemistry*, 50:3942-3944 (1985).
- Veldkamp, W.B., "Binary optics: the optics technology of the 1990s," *CLEO 90*, vol. 7, paper # CMG6 (1990).
- Verlaan-de Vries et al., "A dot-blot screening procedure for mutated ras oncogenes using synthetic oligodeoxynucleotides," *Gene*, 50:313-320 (1986).
- Verpoorte et al., "Three-dimensional micro flow manifolds for miniaturized chemical analysis systems," *J. Microchem. Microeng.*, 4:246-256 (1994).
- Volkmut et al., "DNA electrophoresis in microlithographic arrays," *Nature*, 358:600-602 (1992).
- Voss et al., "The immobilization of oligonucleotides and their hybridization properties," *Biochem. Soc. Transact.*, 16:216-217 (1988).
- Walker et al., "Photolabile Protecting Groups for an Acetylcholine Receptor Ligand. Synthesis and Photochemistry of a New Class of o-Nitrobenzyl Derivatives and their Effects on Receptor Function," *Biochemistry*, 25:1799-1805 (1986).
- Wallace et al., "Hybridization of synthetic oligodeoxyribonucleotides to $\Phi\chi$ 174 DNA: the effect of single base pair mismatch," *Nuc. Acids Res.*, 11(6):3543-3557 (1979).
- Washizu et al., "Handling Biological Cells Using a Fluid Integrated Circuit," *IEEE Transactions Industry Applications*, 26(2):352-358 (1990).
- Werner et al., "Size-Dependent Separation of Proteins Denatured in SDS by Capillary Electrophoresis Using a Replaceable Sieving Matrix," *Anal. Biochem.*, 212:253-258 (1993).
- White et al., "An Evaluation of Confocal Versus Conventional Imaging of Biological Structures by Fluorescence Light Microscopy," *J. Cell Biol.*, 105(1):41-48 (1987).
- Widacki et al., "Biochemical Differences in Qa-2 Antigens Expressed by Qa-2a,6+ and Qa-2a,6- Strains, Evidence for Differential Expression of the o7 and o9 Genes," *Mol. Immunology*, 27(6):559-570 (1990).
- Wilcox et al., "Synthesis of Photolabile 'Precursors' of Amino Acid Neurotransmitters," *J. Org. Chem.*, 55:1585-1589 (1990).

US 6,420,169 B1

Page 9

- Wilding et al., "PCR in a Silicon Microstructure," *Clin. Chem.*, 40(9):1815-1818 (1994).
- Wilding et al., "Manipulation and Flow of Biological Fluids in Straight Channels Micromachined in Silicon," *Clin. Chem.*, 40(1):43-47 (1994).
- Wittman-Liebold, eds., *Methods in Protein Sequence Analysis*, from Proceedings of 7th Int'l Conf., Berlin, Germany, 7/3-8/88, table of contents, pp. xi-xx* (1989).
- Woolley et al., "Ultra-high-speed DNA fragment separations using microfabricated capillary array electrophoresis chips," *PNAS*, 91:11348-11352 (1994).
- Wu et al., "Synthesis and Properties of Adenosine-5'-triphospho- γ -5-(5-sulfonic acid)naphthyl Ethylamide: A Fluorescent Nucleotide Substrate for DNA-Dependent RNA Polymerase from *Escherichia coli*," *Arch. Biochem. Biophys.*, 246(2):564-571 (1986).
- Wu et al., "Laboratory Methods, Direct Analysis of Single Nucleotide Variation in Human DNA and RNA Using In Situ Dot Hybridization," *DNA*, 8(2):135-142 (1989).
- Yamamoto et al., "Features and applications of the laser scanning microscope," *J. Mod. Optics*, 37(11):1691-1701 (1990).
- Yarbrough et al., "Synthesis and Properties of Fluorescent Nucleotide Substrates for DNA-dependent RNA Polymerases," *J. Biol. Chem.*, 254(23):12069-12073 (1979).
- Schuup et al., "Mechanistic Studies of the Photorearrangement of o-Nitrobenzyl Esters," *J. Photochem.*, 36:85-97 (1987).
- Seiler et al., "Planar Glass Chips for Capillary Electrophoresis: Repetitive Sample Injection, Quantitation, and Separation Efficiency," *Anal. Chem.*, 65:1481-1488 (1993).
- Seller et al., "Electroosmotic Pumping and Valveless Control of Fluid Flow within a Manifold of Capillaries on a Glass Chip," *Anal. Chem.*, 66:3485-3491 (1994).
- Semmelhack et al., "Selective Removal of Protecting Groups Using Controlled Potential Electrolysis," *J. Am. Chem. Society*, 94(14):5139-5140 (1972).
- Sheldon et al., "Matrix DNA Hybridization," *Clinical Chemistry*, 39(4):718-719 (1993).
- Shin et al., "Dehydrooligonopeptides. XI. Facile Synthesis of Various Kinds of Dehydrodi- and tripeptides, and Dehydroenkephalins Containing Tyr Residue by Using N-Carboxydehydrotyrosine Anhydride," *Bull. Chem. Soc. Jpn.*, 62:1127-1135 (1989).
- Sim et al., "Use of a cDNA Library for Studies on Evolution and Developmental Expression of the Chorion Multigene Families," *Cell*, 18:1303-1316 (1979).
- Smith et al., "A Novel Method for Delineating Antigenic Determinants: Peptide Synthesis and Radioimmunoassay Using the Same Solid Support," *Immunochemistry*, 14:565-568 (1977).
- Southern et al., "Report on the Sequencing by Hybridization Workshop," *Genomics*, 13:1378-1383 (1992).
- Southern et al., "Oligonucleotide hybridisations on glass supports: a novel linker for oligonucleotide synthesis and hybridization properties of oligonucleotides synthesized in situ," *Nuc. Acids Res.*, 20(7):1679-1684 (1992).
- Southern et al., "Analyzing and Comparing Nucleic Acid Sequences by Hybridization to Arrays of Oligonucleotides: Evaluation Using Experimental Models," *Genomics*, 13:1008-10017 (1992).
- Stemme et al., "A valveless diffuser/nozzle-based fluid pump," *Sensors and Actuators*, A39:159-167 (1993).
- Stryer, L., "DNA Probes and Genes Can be Synthesized by Automated Solid-Phase Methods," from *Biochemistry*, Third Edition, published by W.H. Freeman & Co., (1988).
- Stuber et al., "Synthesis and photolytic cleavage of bovine insulin B22-30 on a nitrobenzoylglycyl-poly (ethylene glycol) support," *Int. J. Peptide Protein Res.*, 22(3):277-283 (1984).
- Sundberg et al., "Spatially-Addressable Immobilization of Macromolecules on Solid Supports," *J. Am. Chem. Soc.*, 117(49):12050-12057 (1995).
- Swedberg, S.A., "Use of non-ionic and zwitterionic surfactants to enhance selectivity in high-performance capillary electrophoresis, An apparent micellar electrokinetic capillary chromatography mechanism," *J. Chromatography*, 503:449-452 (1990).
- Titus et al., "Texas Red, a Hydrophilic, red-emitting fluorophore for use with fluorescein in dual parameter flow microfluorometric and fluorescence microscopic studies," *J. Immunol. Meth.*, 50:193-204 (1982).
- Tkachuk et al., "Detection of bcr-abl Fusion in chronic Myelogenous Leukemia by in situ Hybridization," *Science*, 250:559-562 (90).
- Trzeciak et al., "Synthesis of 'Head-to-Tail' Cyclized Peptides on Solid Support by FMOC Chemistry," *Tetrahed. Letters*, 33(32):4557-4560 (1992).
- Tsien et al., "Control of Cytoplasmic Calcium with Photolabile Tetracarboxylate 2-Nitrobenzhydryl Chelators," *Biophys. J.*, 50:843-853 (1986).
- Tsutsumi et al., "Expression of L- and M-Type Pyruvate Kinase in Human Tissues," *Genomics*, 2:86-89 (1988).
- Turchinskii et al., "Multiple Hybridization in Genome Analysis, Reaction of Diamines and Bisulfate with Cytosine for Introduction of Nonradioactive labels Into DNA," *Molecular Biology*, 22:1229-1235 (1988).
- Turner et al., "Photochemical Activation of Acylated α -Thrombin," *J. Am. Chem. Soc.*, 109:1274-1275 (1987).
- Urdea et al., "A novel method for the rapid detection of specific nucleotide sequences in crude biological samples without blotting or radioactivity; application to the analysis of hepatitis B virus in human serum," *Gene*, 61:253-264 (1987).
- Yosomiya et al., "Performance, Glass fiber Having Isocyanate Group on the Surface. Preparation and Reaction with Amino Acid," *Polymer Bulletin*, 12:41-48 (1984).
- Young, W.S., "Simultaneous Use of Digoxigenin- and Radio-labeled Oligodeoxyribonucleotide Probes for Hybridization Histochemistry," *Neuropeptides*, 13:271-275 (1989).
- Yue et al., "Miniature Field-Flow Fractionation System for Analysis of Blood Cells," *Clin. Chem.*, 40(9):1810-1814 (1994).
- Zehavi et al., "Light-Sensitive Glycosides. I. 6-Nitroveratryl β -D-Glucopyranoside and 2-Nitrobenzyl β -D-Glucopyranoside," *J. Org. Chem.*, 37(14):2281-2285 (1972).
- Zengerle et al., "Transient measurements on miniaturized diaphragm pumps in microfluid systems," *Sensors and Actuators*, A46-47:557-561 (1995).
- Barany, F., "Genetic disease detection and DNA amplification using cloned thermostable ligase," *PNAS*, 81:1991-1995 (1991).
- Chetverin et al., "Oligonucleotide Arrays: New Concepts and Possibilities," *Biotechnology*, 12:1093-1099 (1994).
- Church et al., "Multiplex DNA sequencing," *Science*, 240:185-188 (1988).

US 6,420,169 B1

Page 10

- Church et al., "Genomic sequencing," *PNAS*, 81:1991–1995 (1984).
- Coulson et al., "Toward a physical map of the genome of the nematode *Caenorhabditis elegans*," *PNAS*, 83:7821–7825 (1986).
- Dower et al., "The Search for Molecular Diversity (II): Recombinant and Synthetic Randomized Peptide Libraries," *Ann. Rep. Med. Chem.*, 26:271–280 (1991).
- Drmanac et al., "An Algorithm for the DNA Sequence Generation from k-Tuple World Contents of the Minimal Number of Random Fragments," *J. Biomol. Struct. Dyn.*, 8(5):1085–1102 (1991).
- Elder, J.K., "Analysis of DNA Oligonucleotide Hybridization Data by Maximum Entropy," in *Maximum Entropy and Bayesian Methods*, eds. Mohammad-Djafari and Demoment, Kluwer, Dordrecht, pp. 363–371 (1992).
- Feinberg et al., ADDENDUM to "A technique for Radiolabeling DNA Restriction Endonuclease Fragments to High Specific Activity," *Anal. Biochem.*, 137:266–267 (1984).
- Hodgson et al., "Hybridization probe size control: optimized 'oligolabelling'," *Nuc. Acids Res.*, 15(15):6295 (1987).
- Khrapko et al., "A method for DNA sequencing by hybridization with oligonucleotide matrix," *DNA Seq. Map.*, 1:375–388 (1991).
- Lander et al., "Genomic Mapping by Fingerprinting Random Clones: A Mathematical Analysis," *Genomics*, 2:231–239 (1988).
- Little, P., "Clone maps made simple," *Nature*, 346:611–612 (1990).
- Luo, J. et al., "Improving the fidelity of *Thermus thermophilus* DNA ligase," *Nuc. Acids Res.*, 24(14):3071–3078 (1996).
- Olson et al., "Random-clone strategy for genomic restriction mapping in yeast," *PNAS*, 83:7826–7830 (1986).
- Pevzner, P.A., "DNA Physical Mapping and Alternating Eulerian Cycles in Colored Grapes," *Algorithmica*, 13(1–2):77–105 (1995).
- Pevzner et al., "Multiple Filtration and Approximate Pattern Matching," *Algorithmica*, 13(1–2):135–154 (1995).
- Pevzner et al., "Generalized Sequence Alignment and Duality," *Adv. Applied Math.*, 14:139–XXX (1993).
- Pfeifer et al., "Genomic Sequencing and Methylation Analysis by Ligation Mediated PCR," *Science*, 246:810–813 (1989).
- Schena et al., "Parallel human genome analysis: Microarray-based expression monitoring of 1000 genes," *PNAS*, 93:10614–10619 (1996).
- Seed, B., "Diazotizable arylamine cellulose papers for the coupling and hybridization of nucleic acids," *Nuc. Acids Res.*, 10(5):1799–1810 (1982).
- Sofia, M.J., "Carbohydrate-based combinatorial libraries," *Molecular Diversity*, 3:75–94 (1998).
- Wallace et al., "The use of synthetic oligonucleotides as hybridization probes. II. Hybridization of oligonucleotides of mixed sequence to rabbit β -globin DNA," *Nuc. Acids Res.*, 9(4):879 (1981).
- Wiedmann, M. et al., "Ligase Chain Reaction (LCR)—Overview and Applications," *PCR Meth. Appl.*, 3(4):S51–S64 (1994).
- Wood et al., "Base composition-independent hybridization in tetramethylammonium chloride: A method for oligonucleotide screening of highly complex gene libraries," *PNAS*, 82:1585–1588 (1985).

* cited by examiner

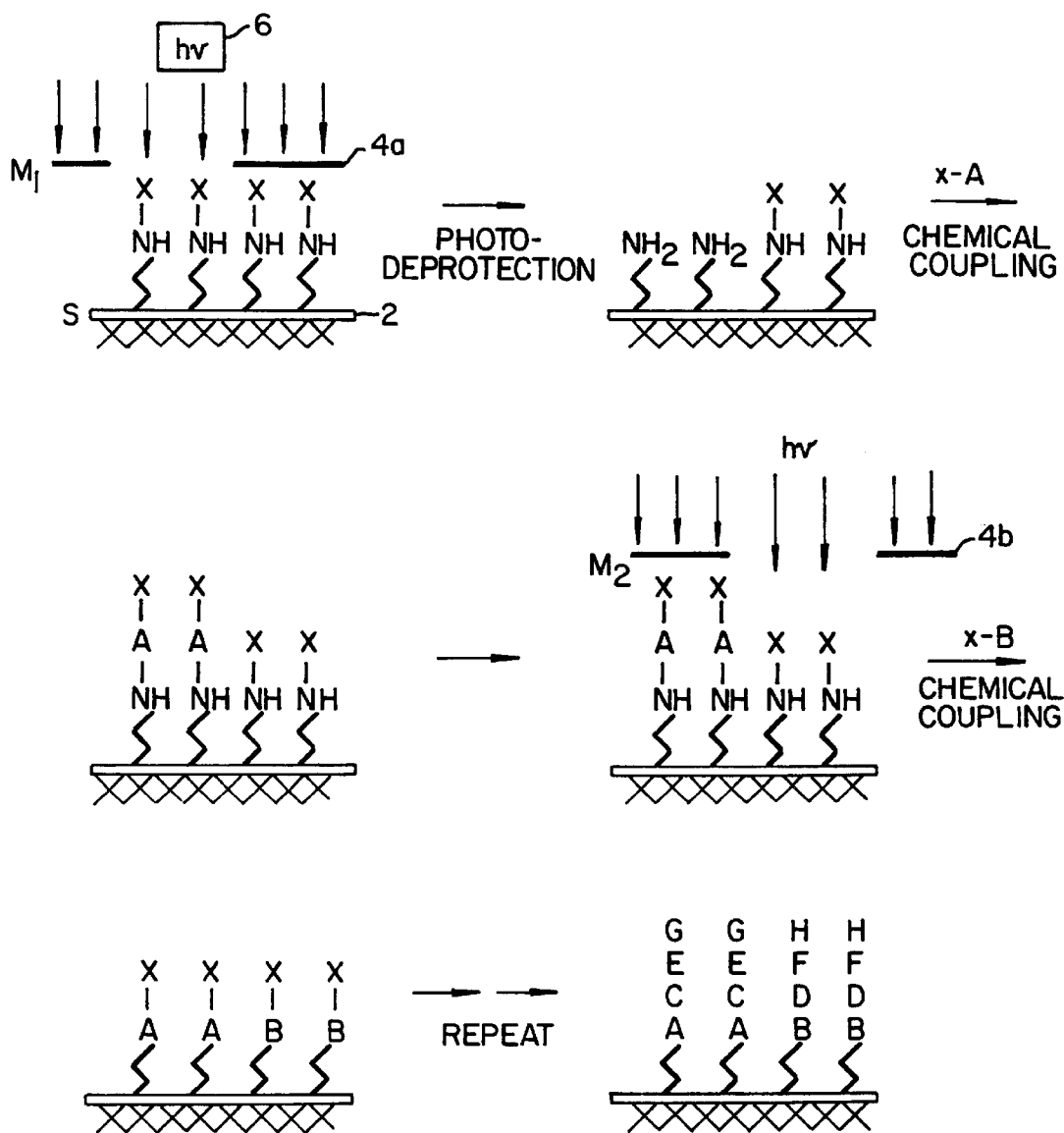


FIG. 1.

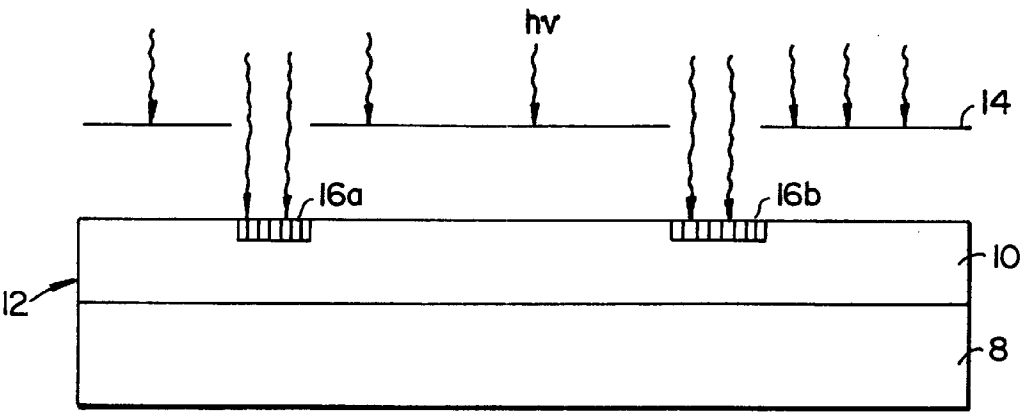


FIG. 2.

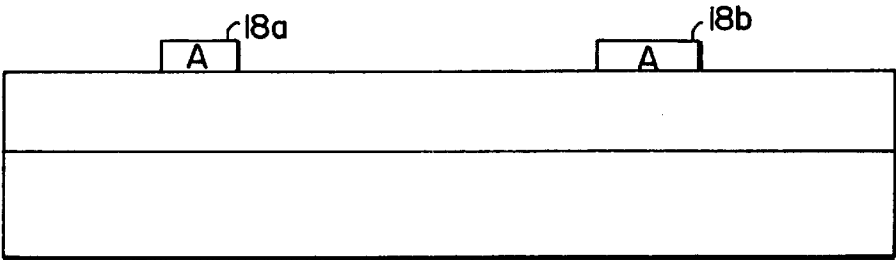


FIG. 3.

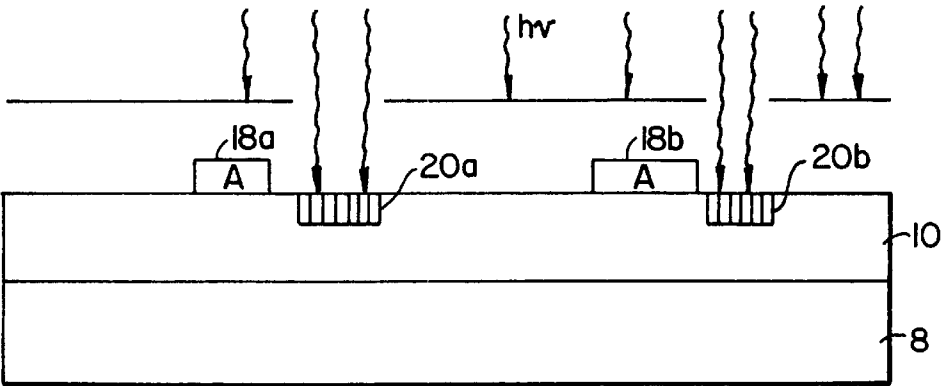


FIG. 4.

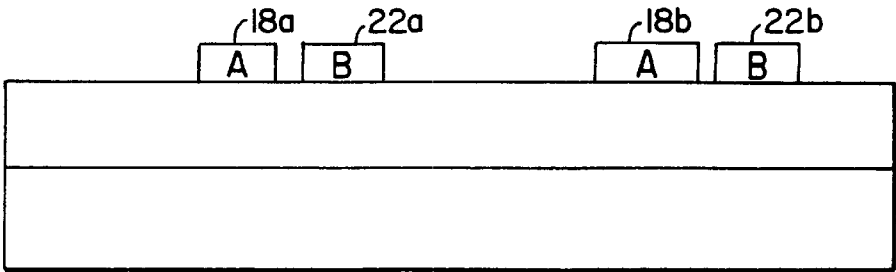


FIG. 5.

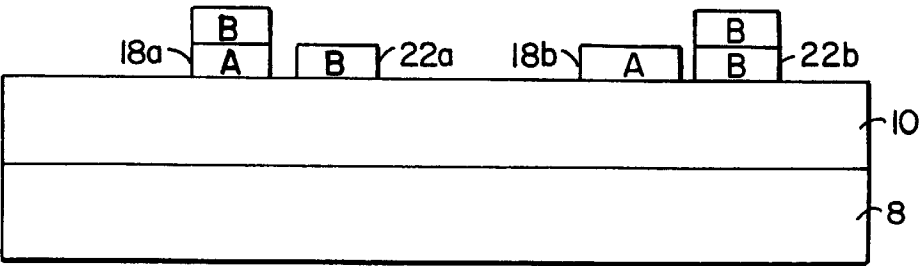


FIG. 6.

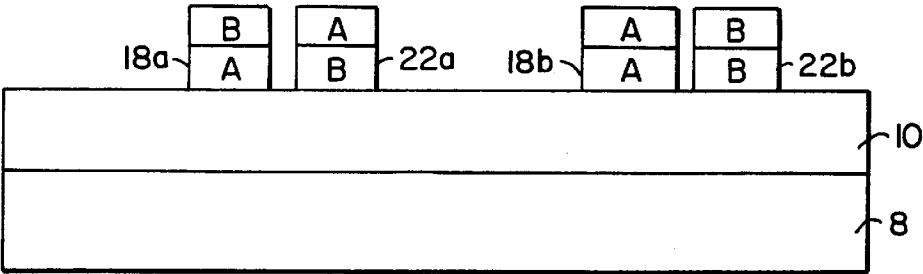


FIG. 7.

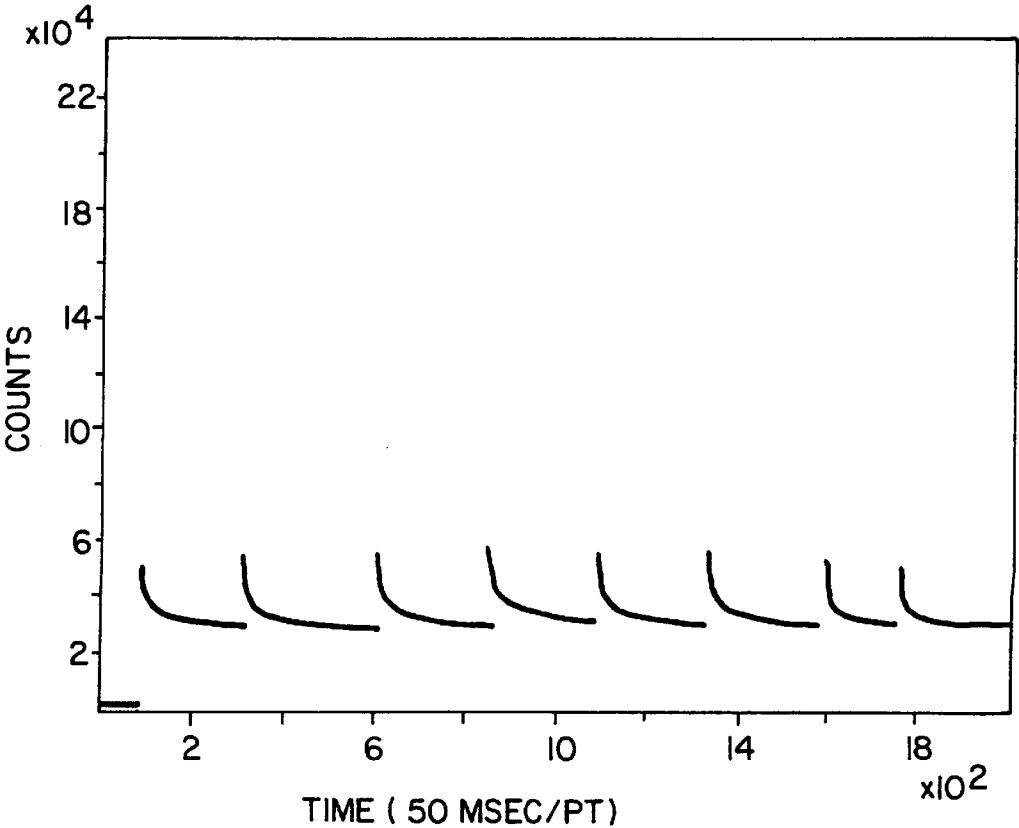


FIG. 8A.

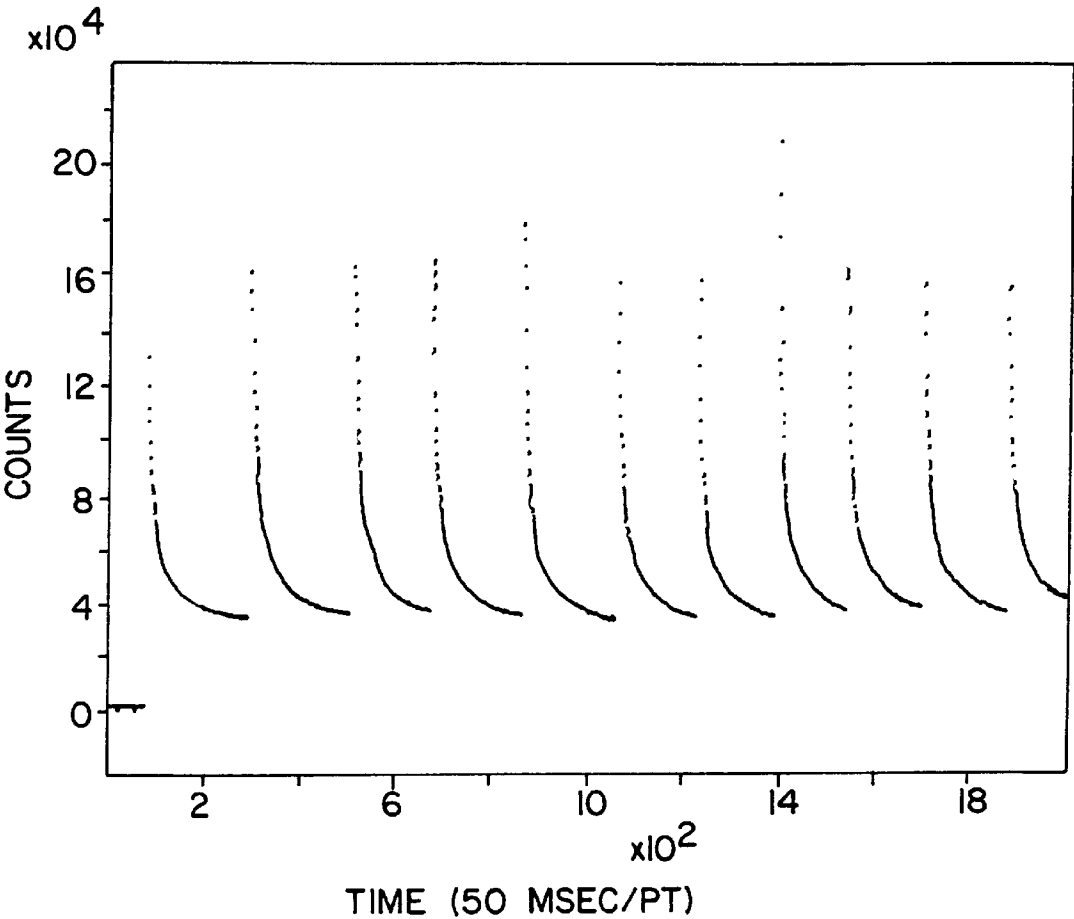


FIG. 8B.

U.S. Patent

Jul. 16, 2002

Sheet 5 of 42

US 6,420,169 B1

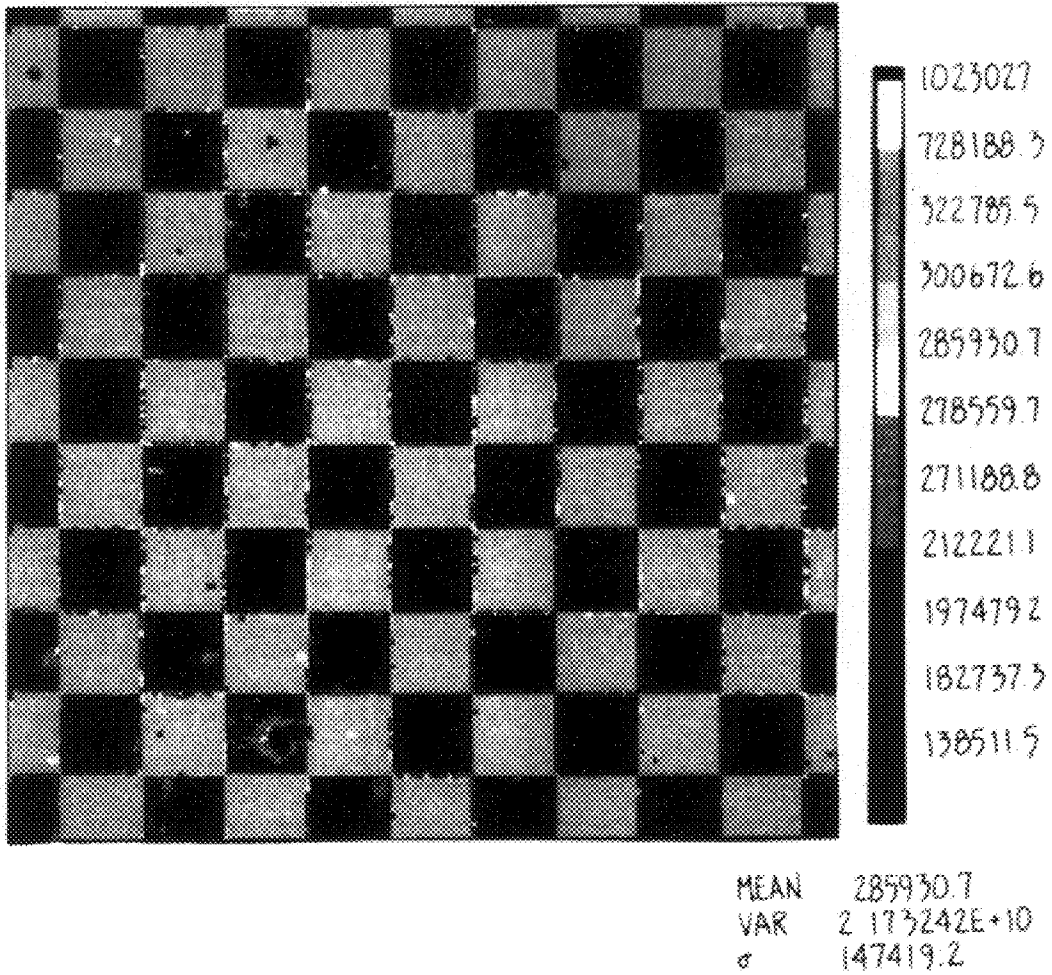


FIG. 9A.

U.S. Patent

Jul. 16, 2002

Sheet 6 of 42

US 6,420,169 B1

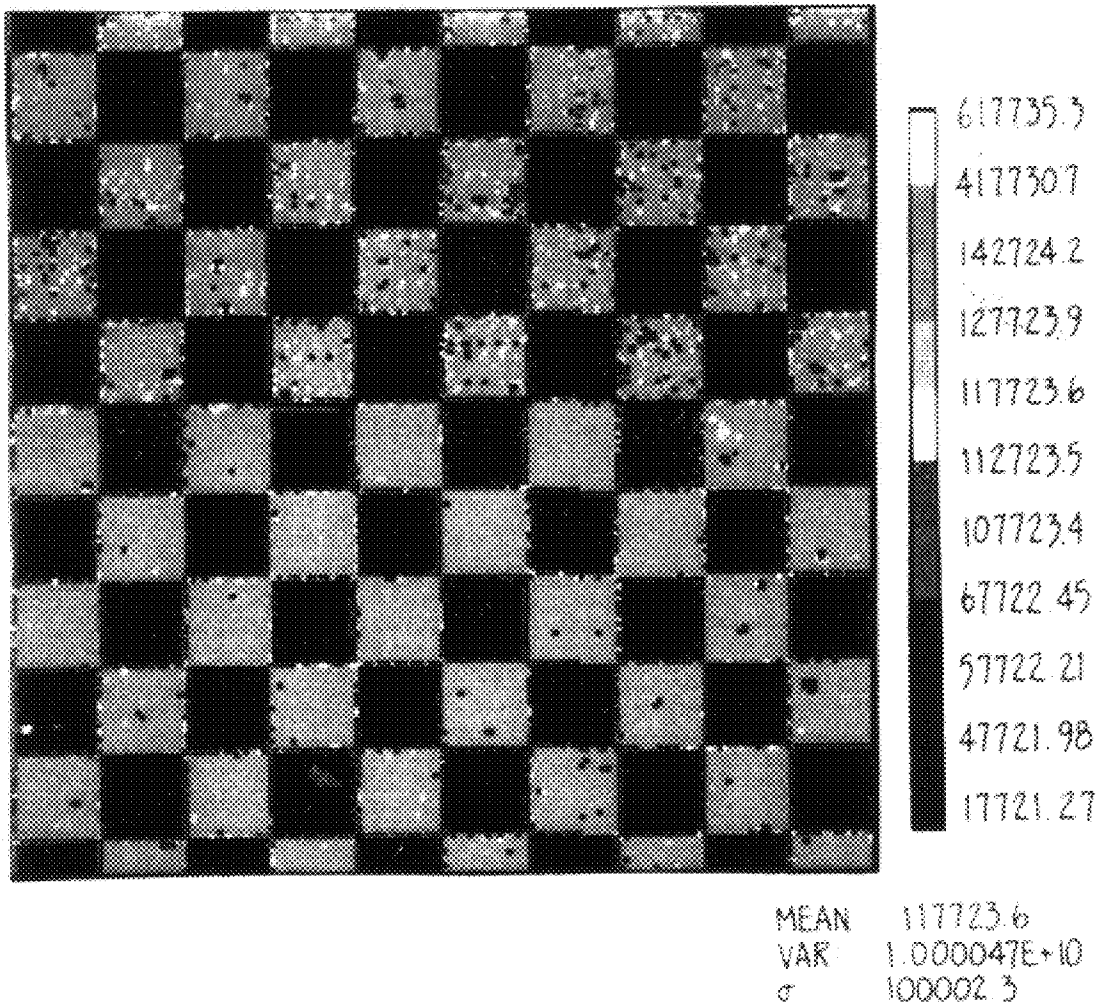


FIG. 9B.

U.S. Patent

Jul. 16, 2002

Sheet 7 of 42

US 6,420,169 B1

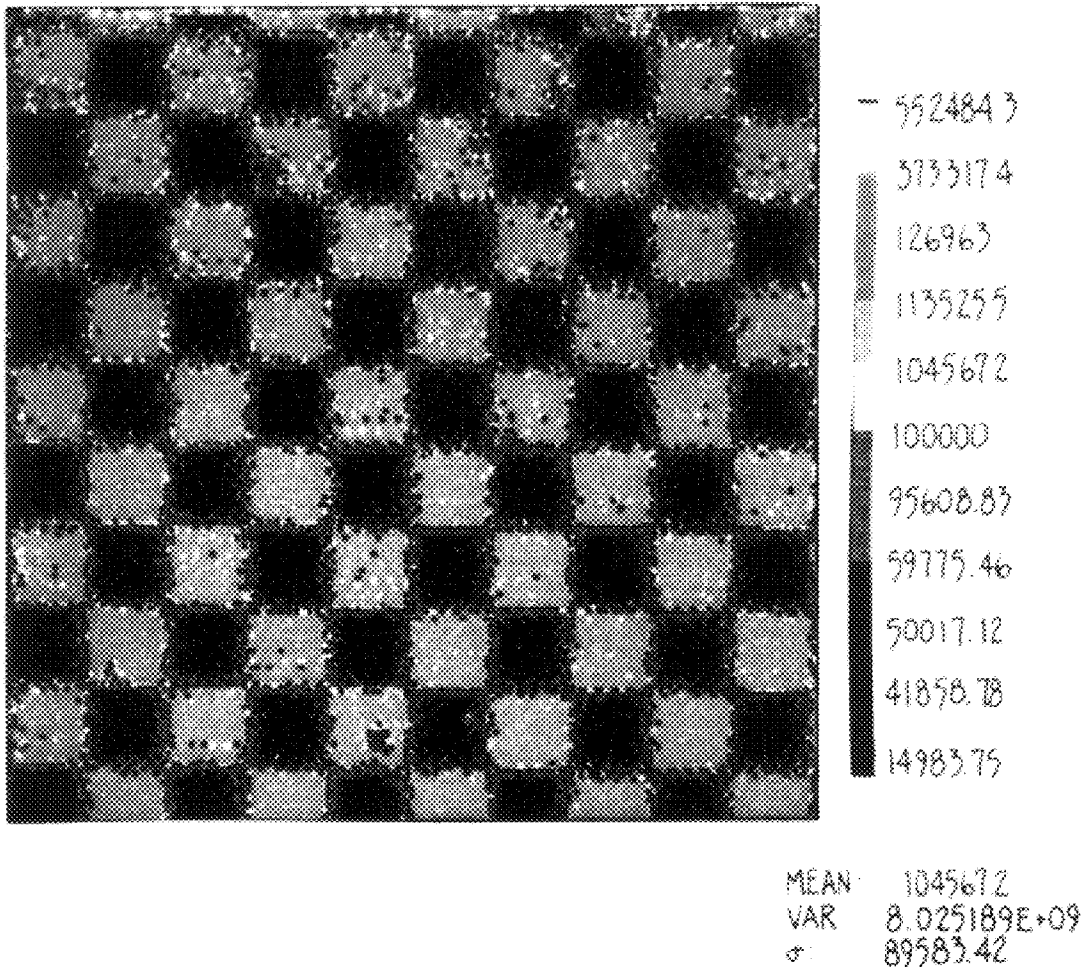


FIG. 9C.

U.S. Patent

Jul. 16, 2002

Sheet 8 of 42

US 6,420,169 B1

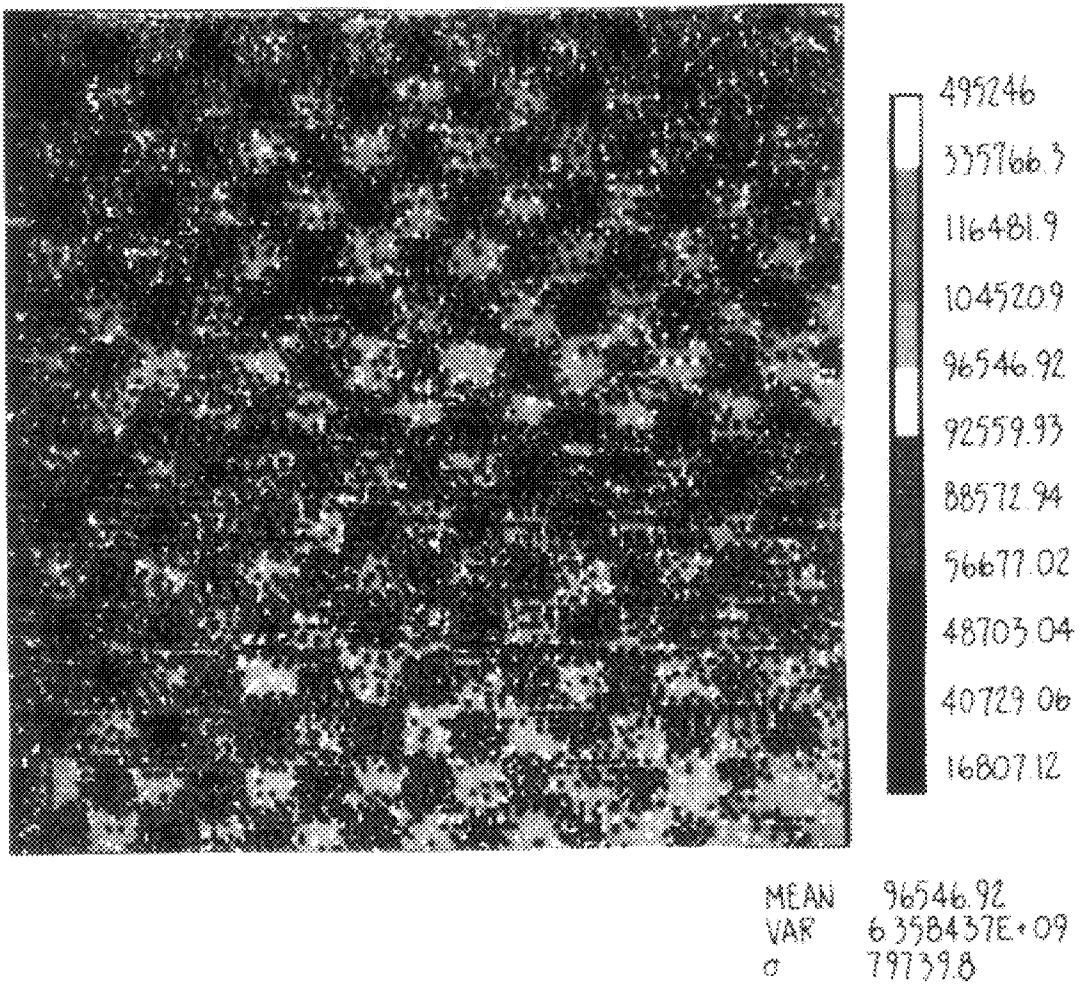


FIG. 9D.

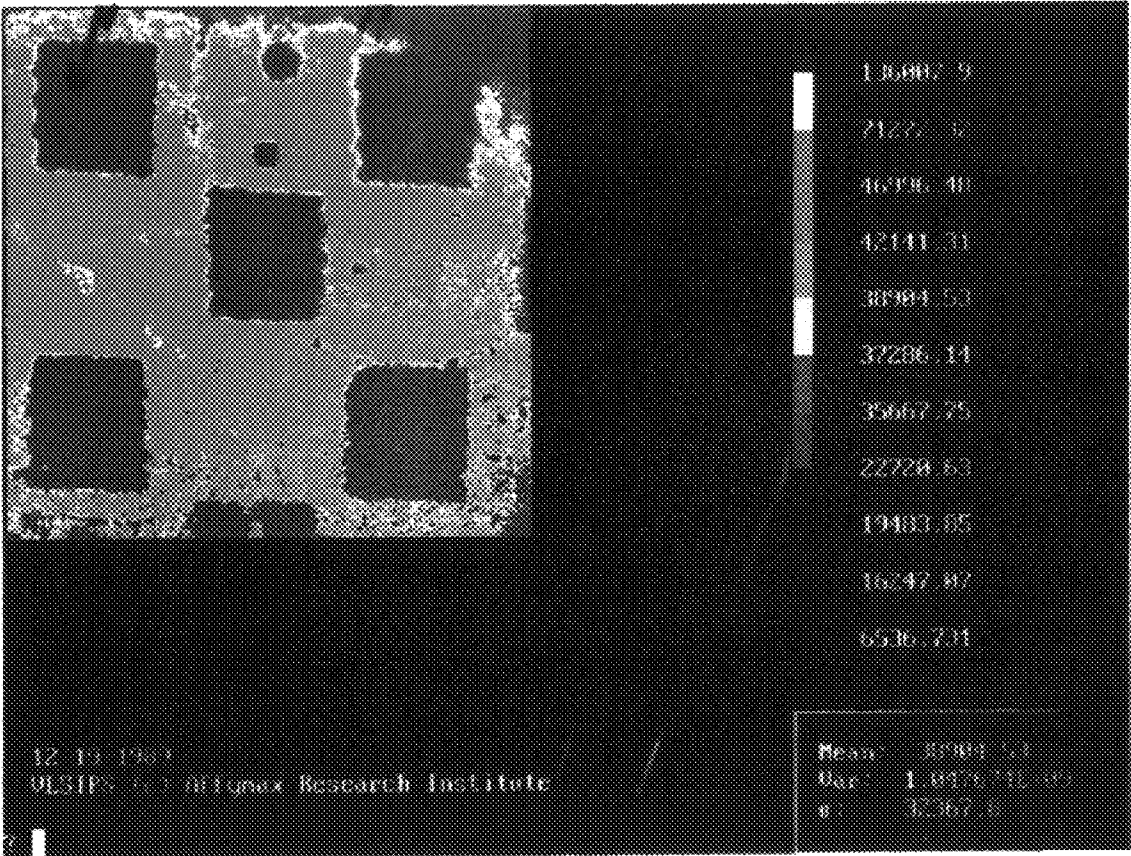


FIG. 10.

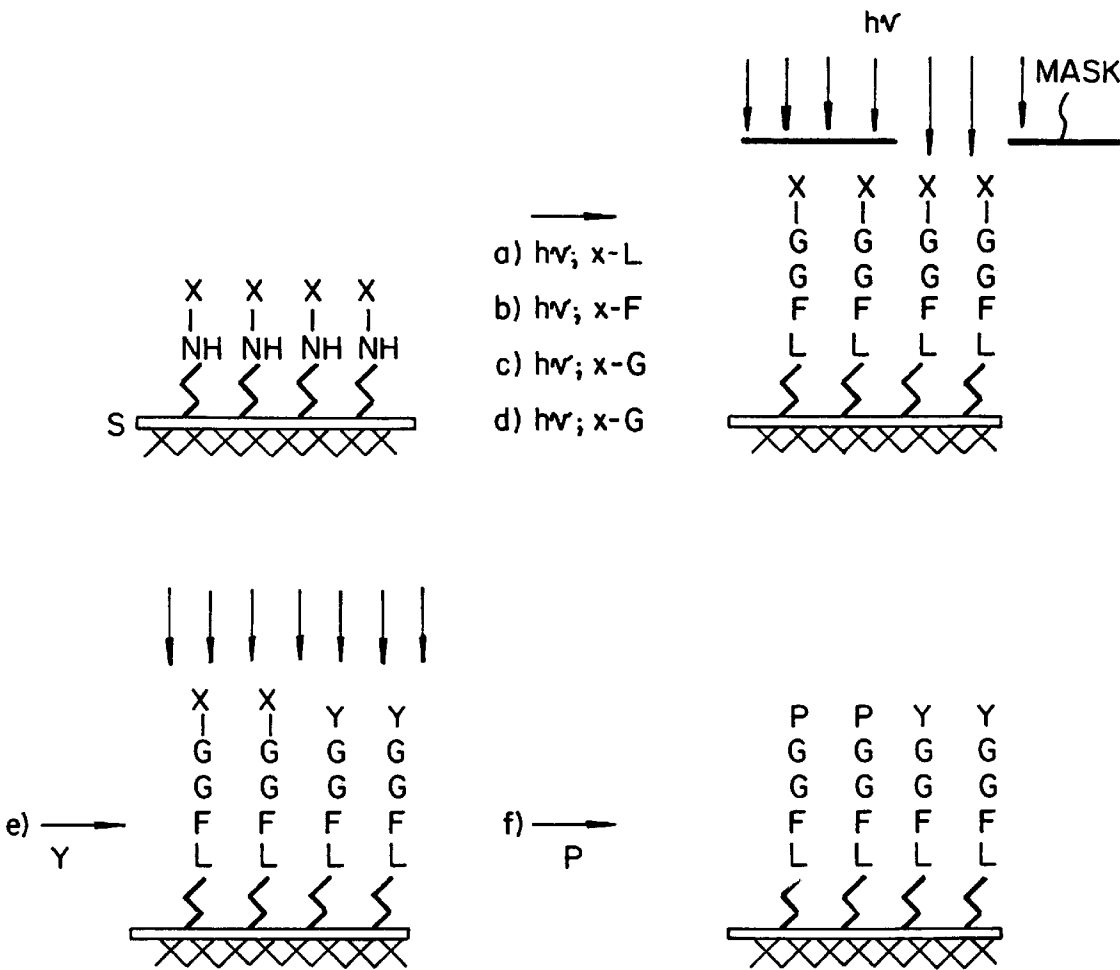


FIG. 11.

U.S. Patent

Jul. 16, 2002

Sheet 11 of 42

US 6,420,169 B1

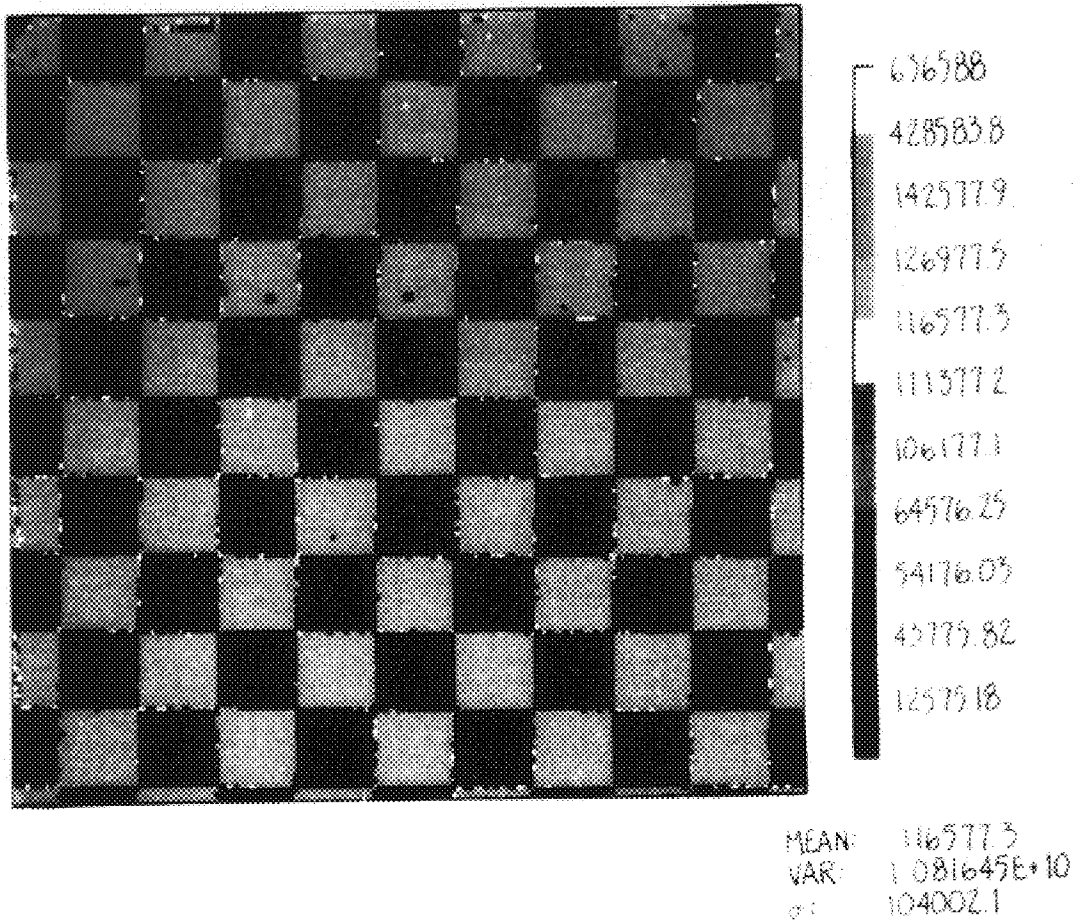


FIG. 12.

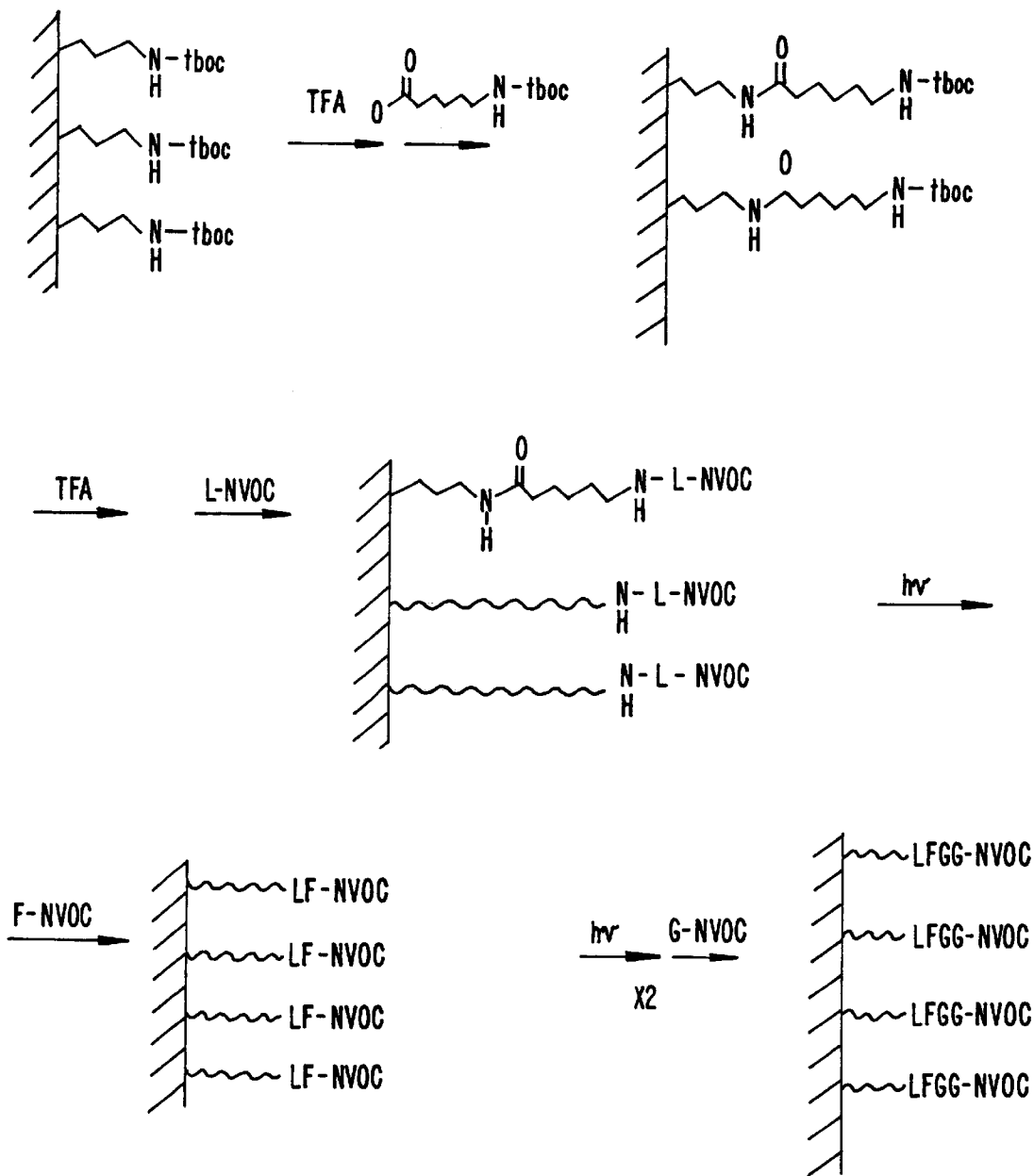


FIG. 13A.

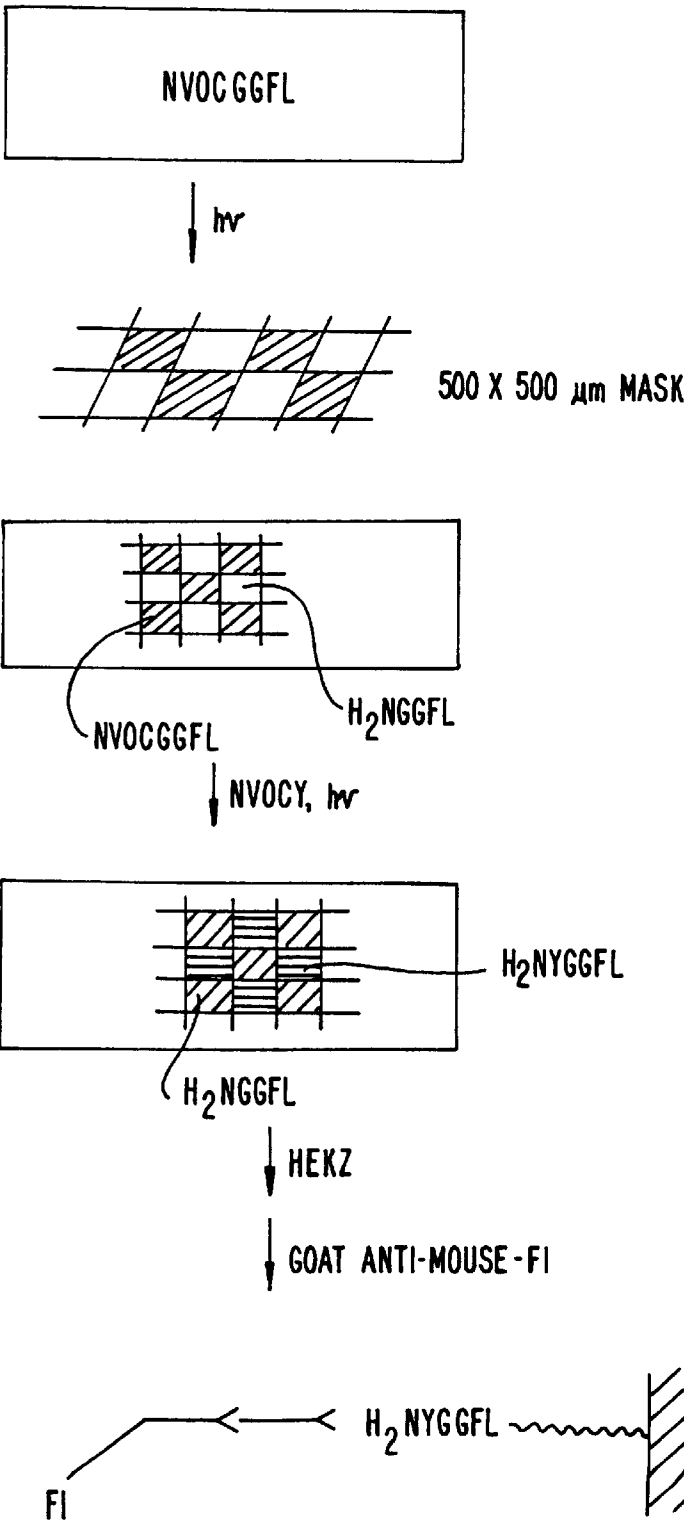


FIG. 13B.

U.S. Patent

Jul. 16, 2002

Sheet 14 of 42

US 6,420,169 B1

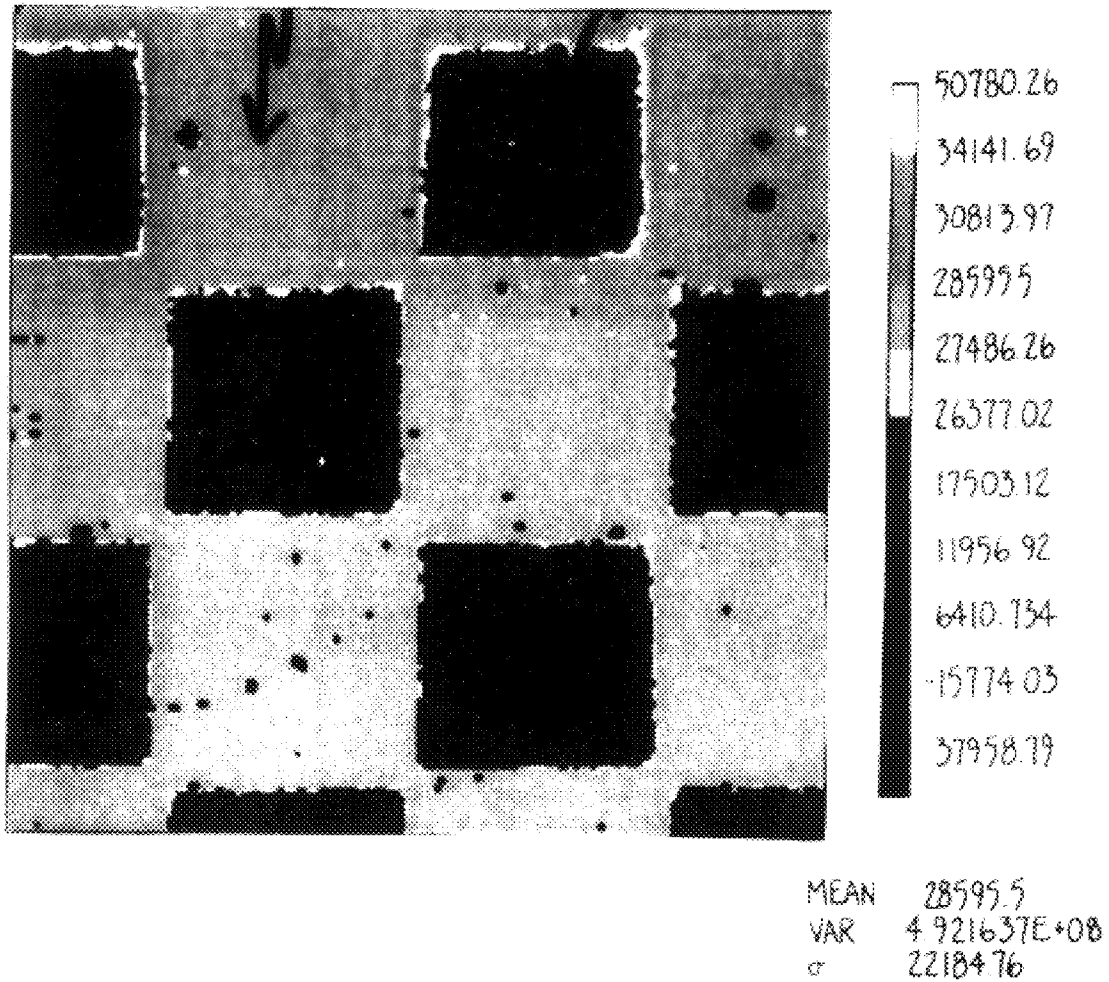


FIG. 13C.

U.S. Patent

Jul. 16, 2002

Sheet 15 of 42

US 6,420,169 B1

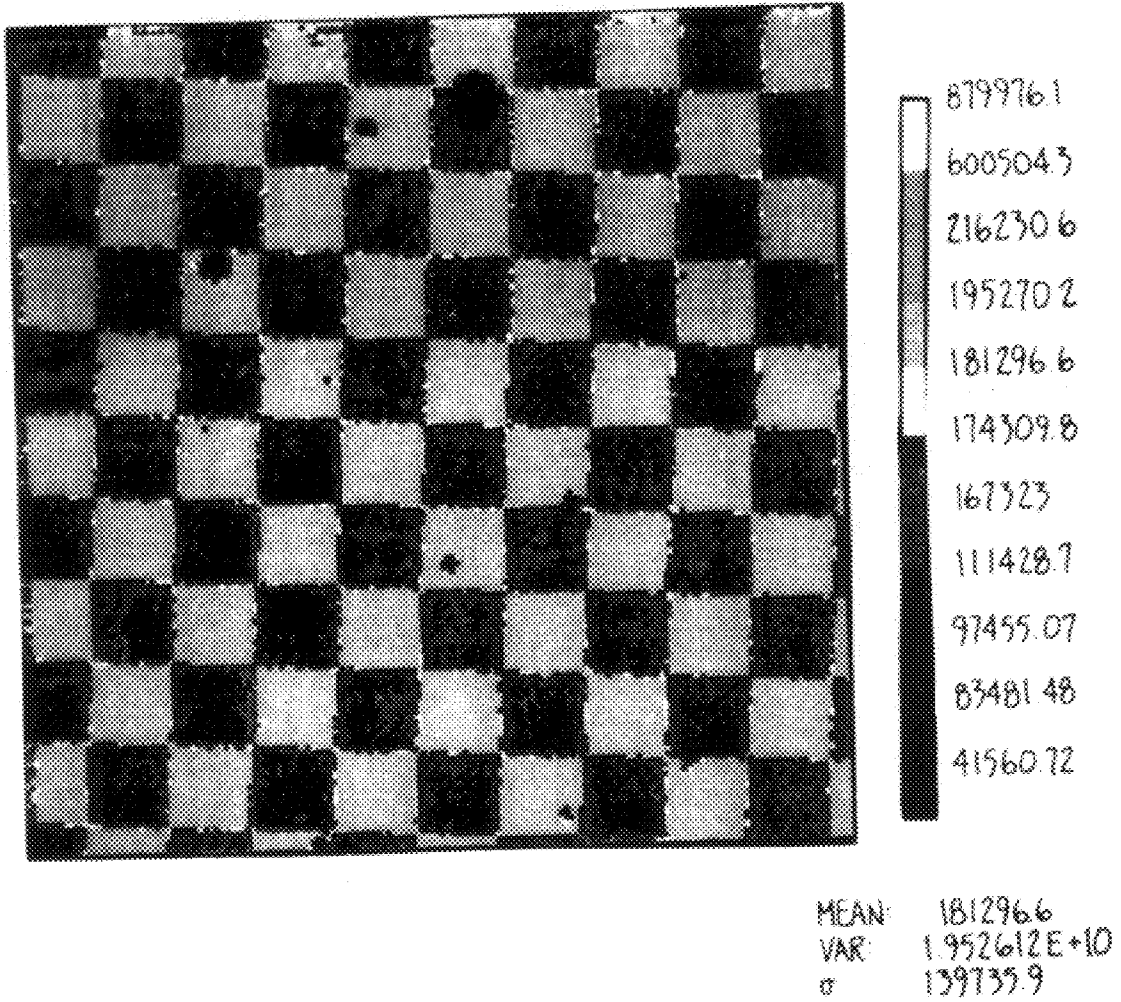


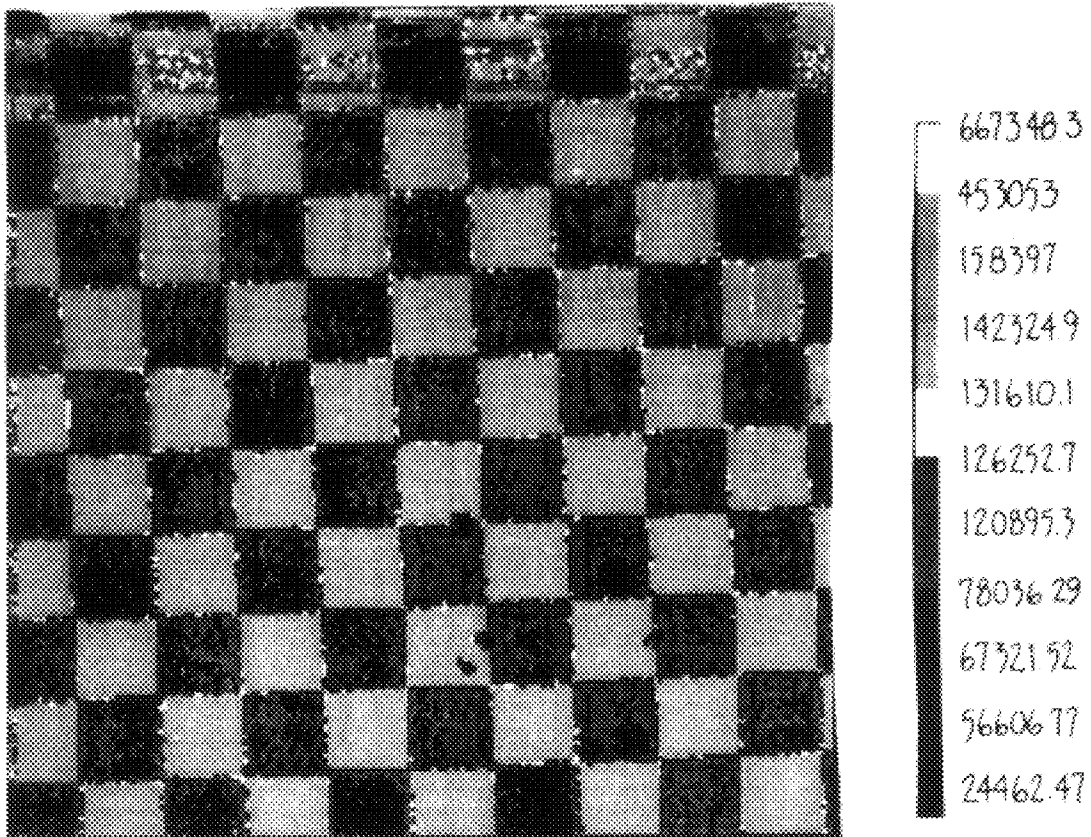
FIG._13D.

U.S. Patent

Jul. 16, 2002

Sheet 16 of 42

US 6,420,169 B1



MEAN 131610.1
VAR 1.148062E+10
 σ 107147.6

FIG._14.

P	A	S	G	
<u>L</u> PGFL	<u>L</u> AGFL	<u>L</u> SGFL	<u>L</u> GGFL	L
<u>F</u> PGFL	<u>F</u> AGFL	<u>F</u> SGFL	<u>F</u> GGFL	F
<u>W</u> PGFL	<u>W</u> AGFL	<u>W</u> SGFL	<u>W</u> GGFL	W
<u>Y</u> PGFL	<u>Y</u> AGFL	<u>Y</u> SGFL	<u>Y</u> GGFL	Y

L SET

FIG. 15A.

p	a	s	G	
<u>Y</u> pGFL	<u>Y</u> aGFL	<u>Y</u> sGFL	<u>Y</u> GGFL	Y
<u>f</u> pGFL	<u>f</u> aGFL	<u>f</u> sGFL	<u>f</u> GGFL	f
<u>w</u> pGFL	<u>w</u> aGFL	<u>w</u> sGFL	<u>w</u> GGFL	w
<u>y</u> pGFL	<u>y</u> aGFL	<u>y</u> sGFL	<u>y</u> GGFL	y

D SET

FIG. 15B.

U.S. Patent

Jul. 16, 2002

Sheet 18 of 42

US 6,420,169 B1

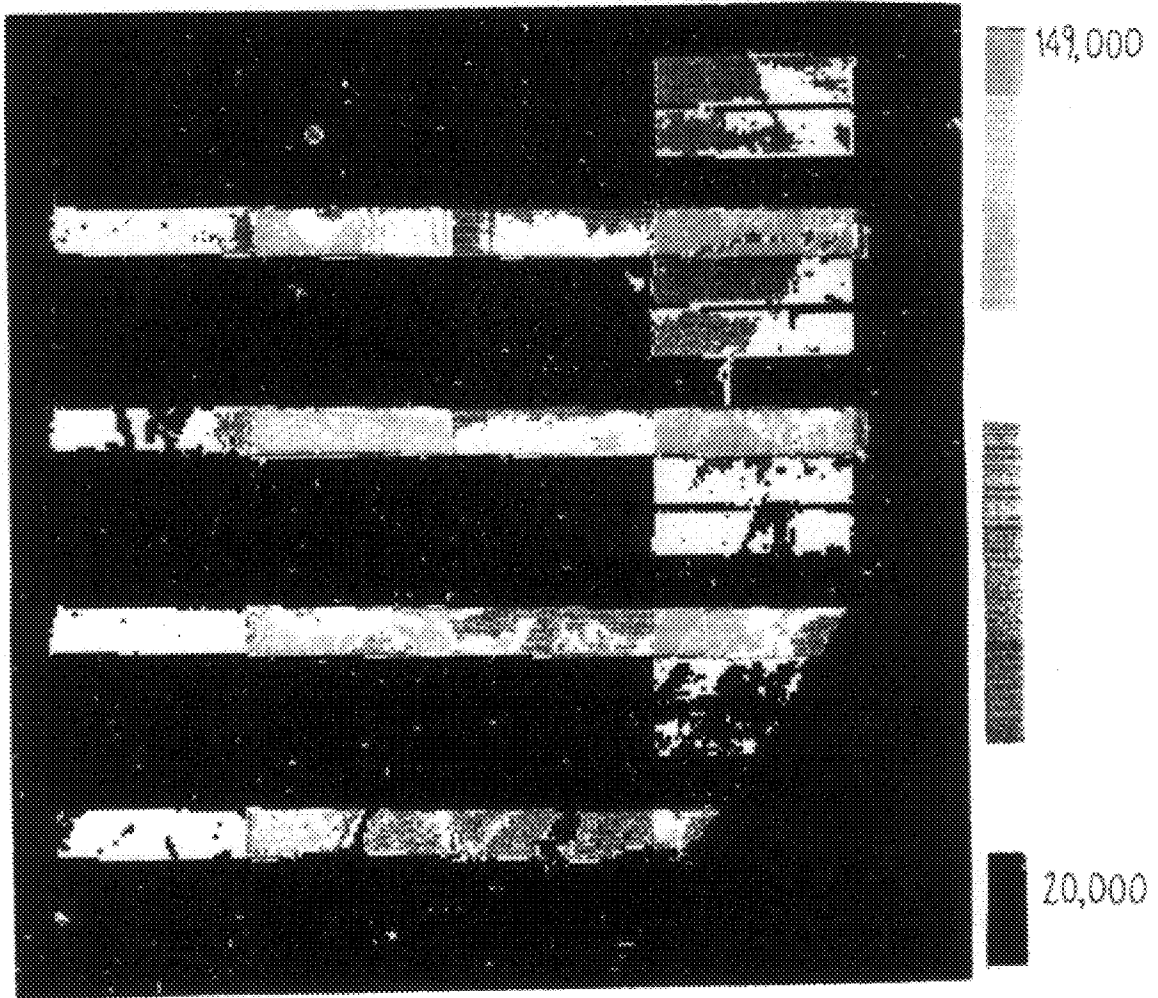


FIG. 16.

U.S. Patent

Jul. 16, 2002

Sheet 19 of 42

US 6,420,169 B1

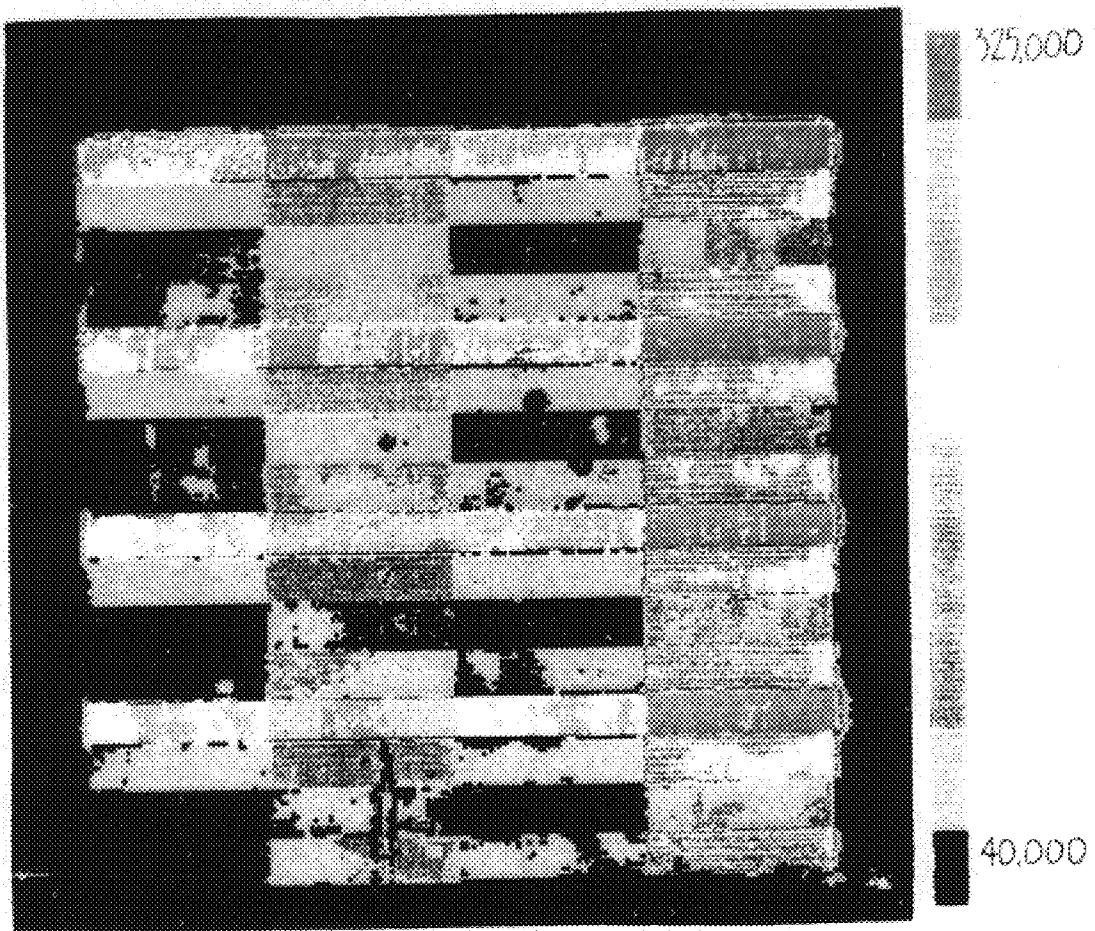


FIG. 17.

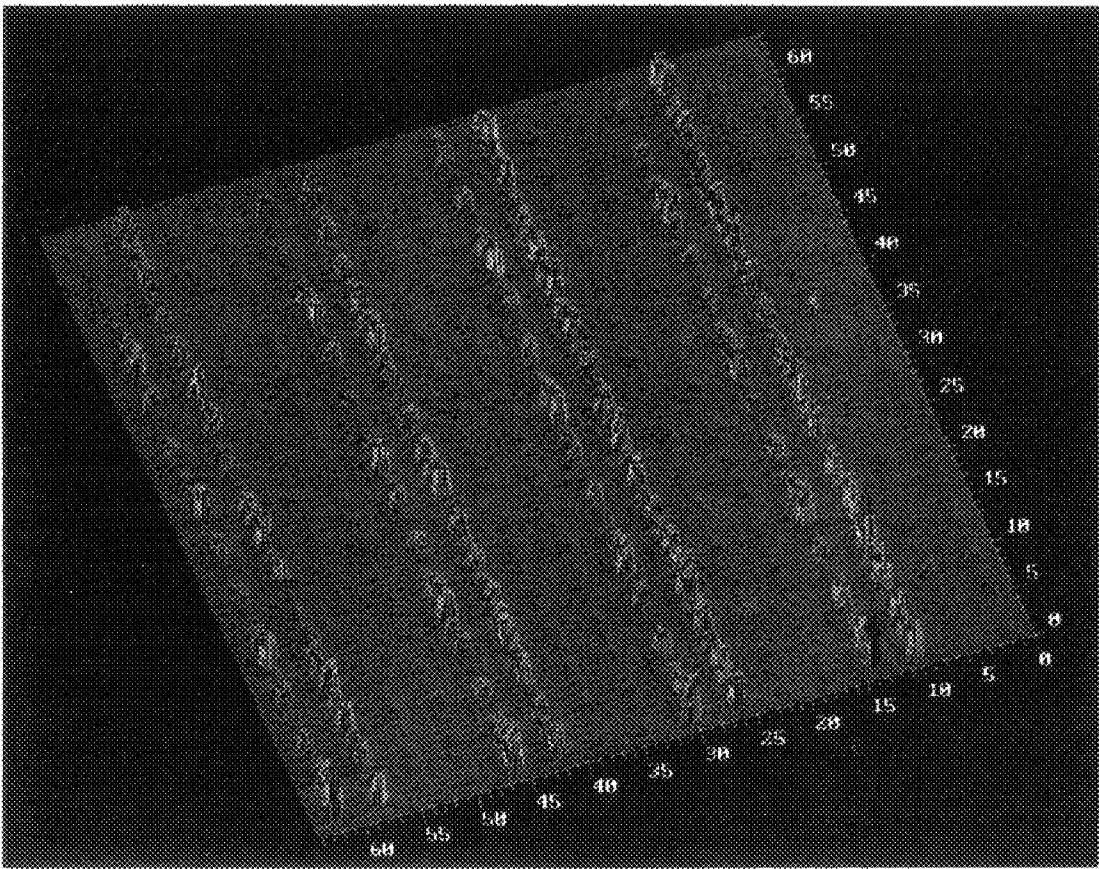


FIG. 18.

U.S. Patent

Jul. 16, 2002

Sheet 21 of 42

US 6,420,169 B1

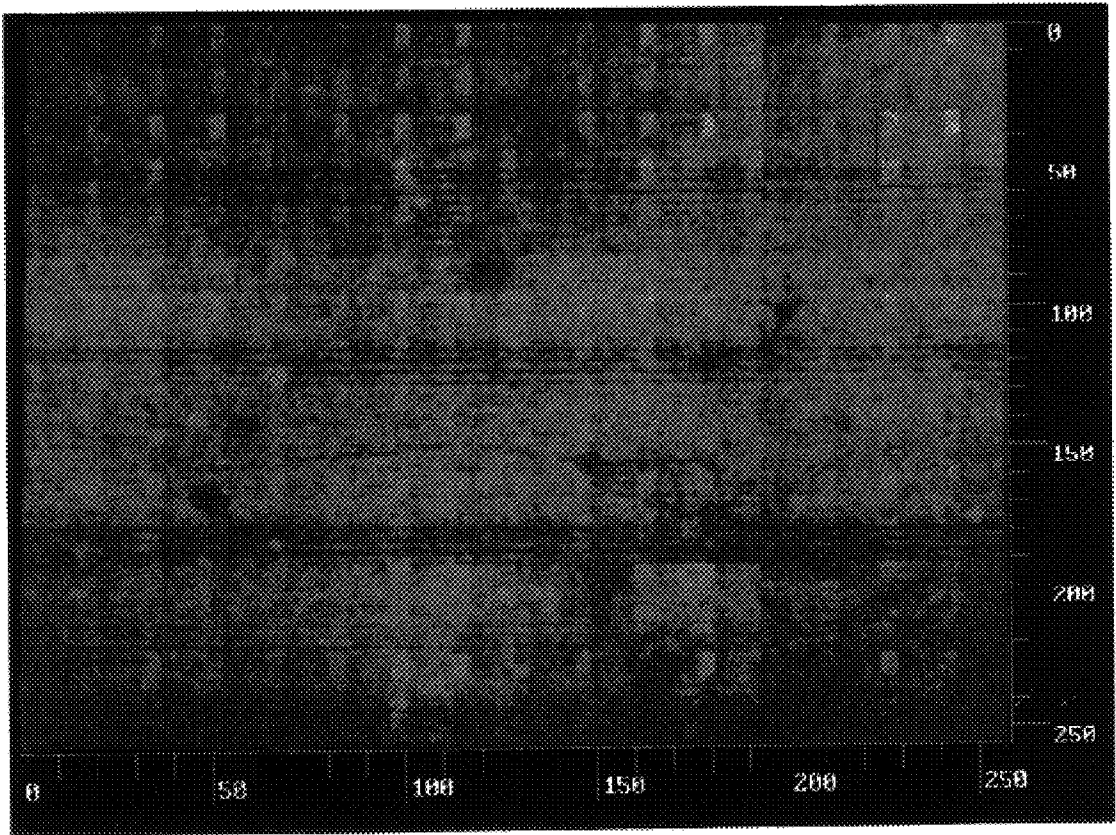


FIG. 19.

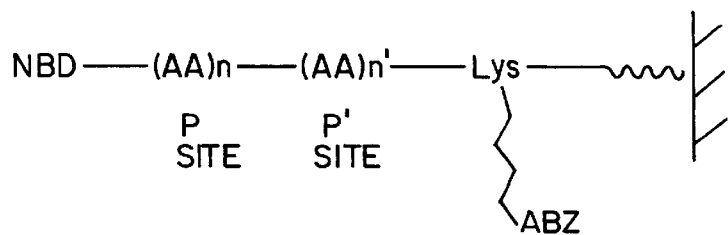


FIG. 20A.

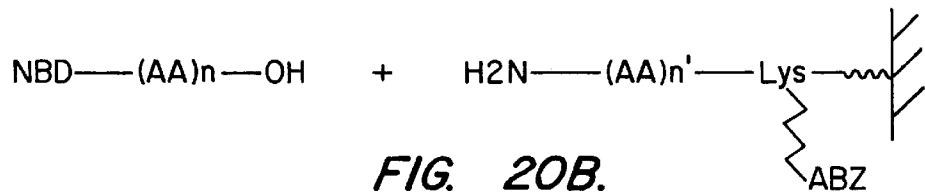


FIG. 20B.

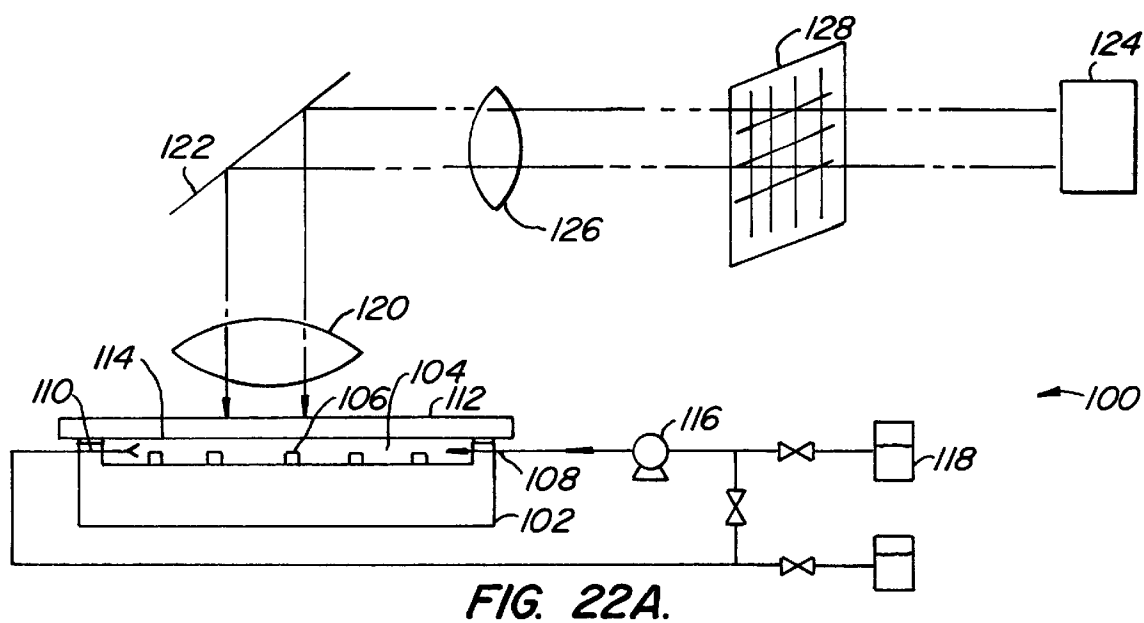


FIG. 22A.

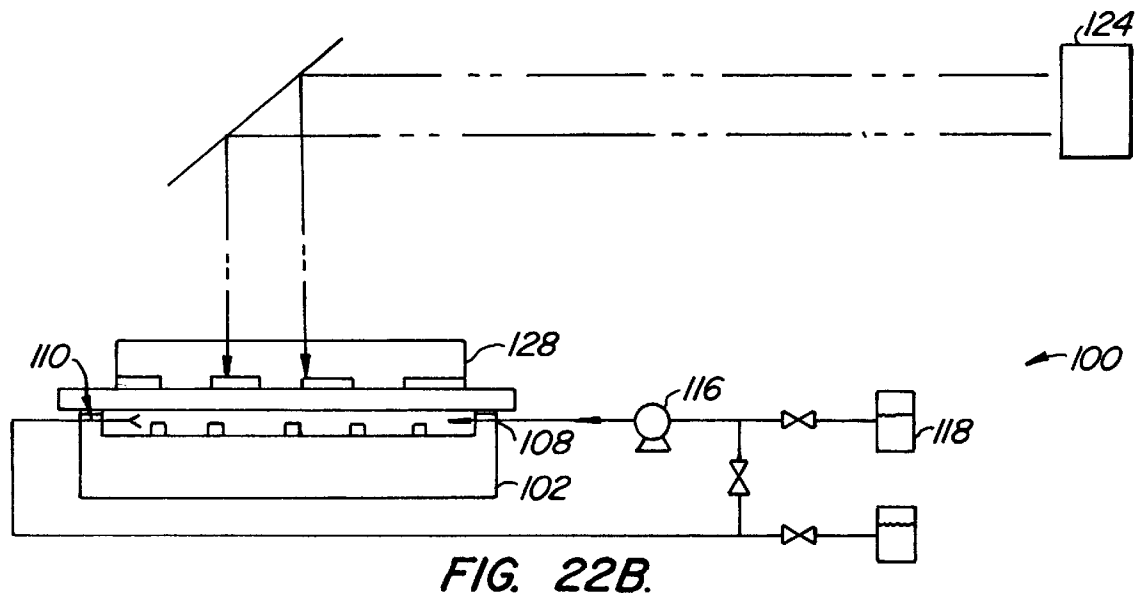


FIG. 22B.

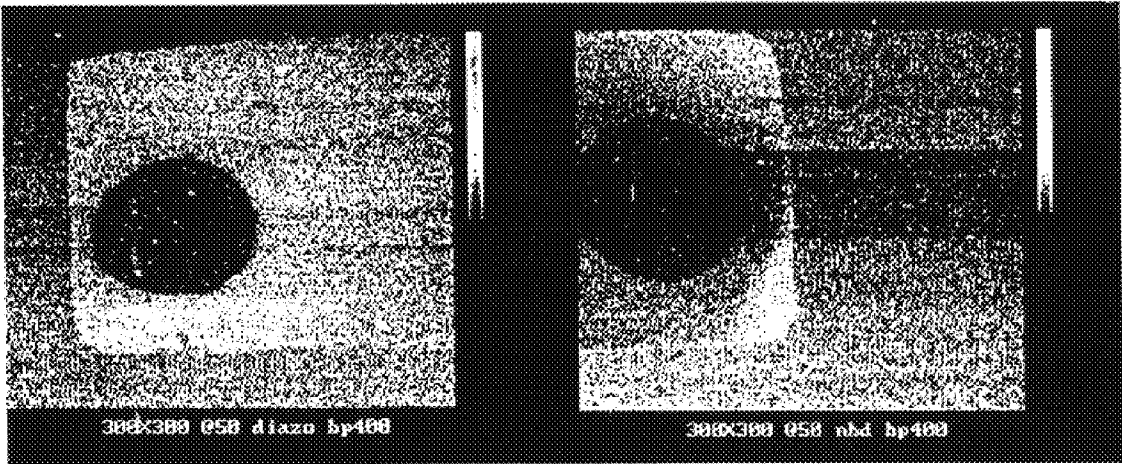


FIG. 21A.

FIG. 21B.

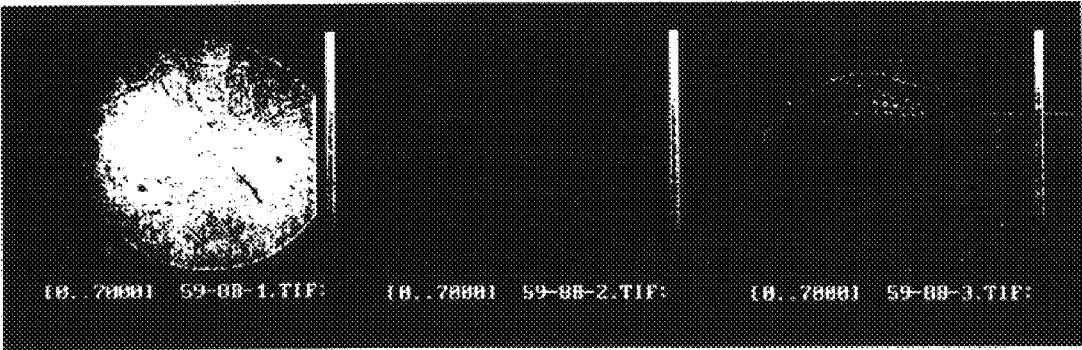


FIG. 39A.

FIG. 39B.

FIG. 39C.

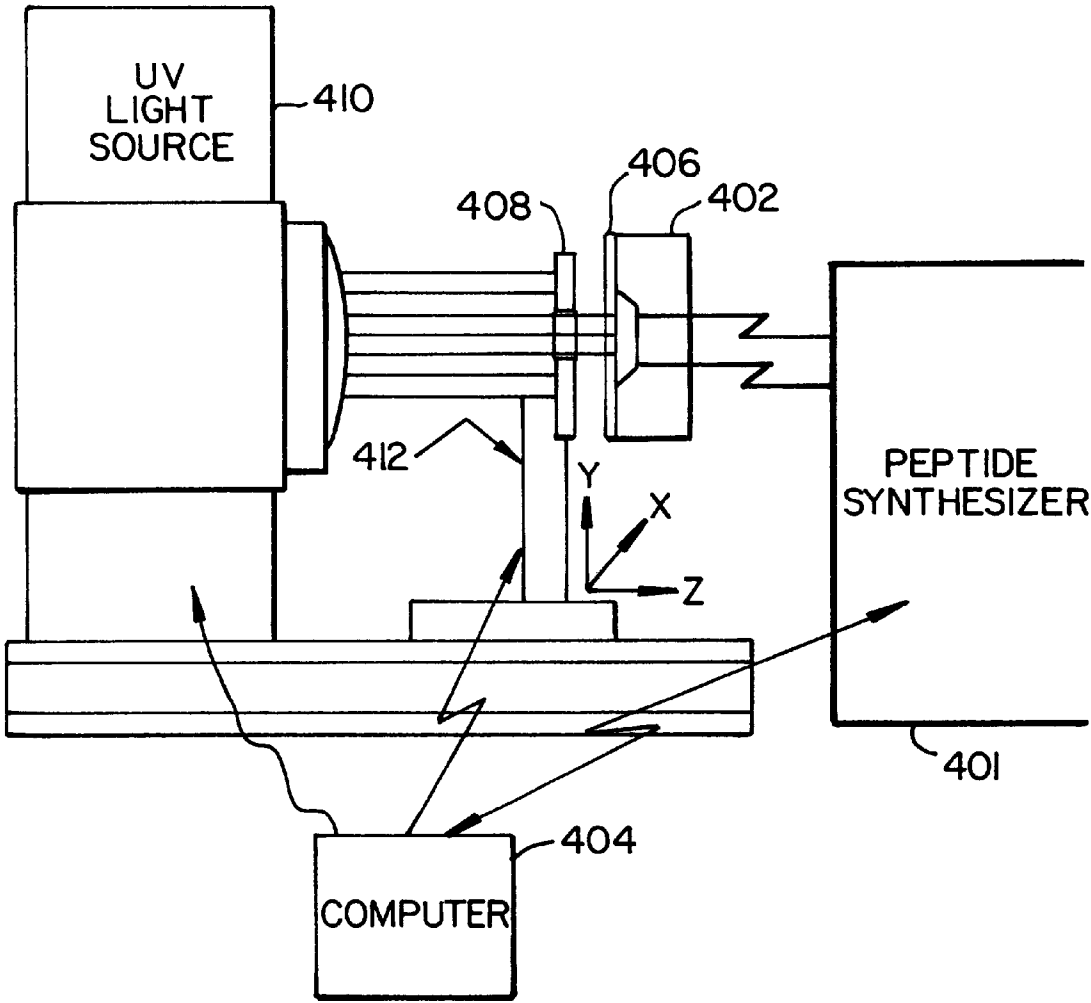


FIG. 23.

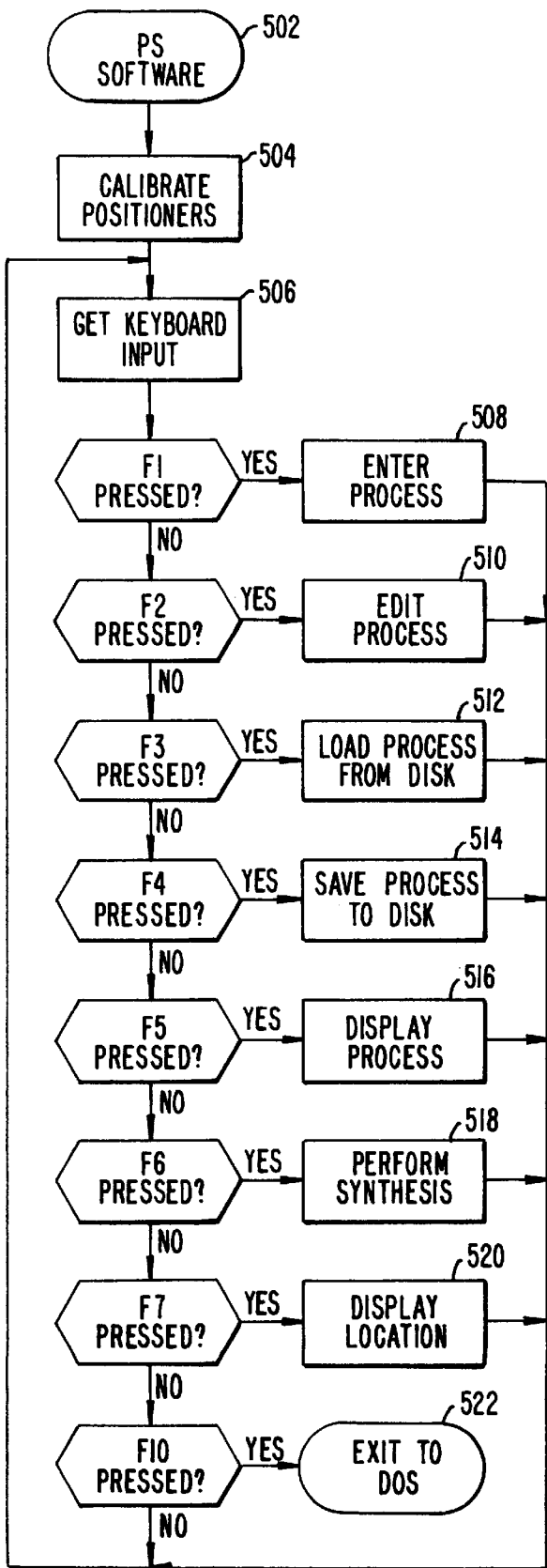


FIG. 24A.

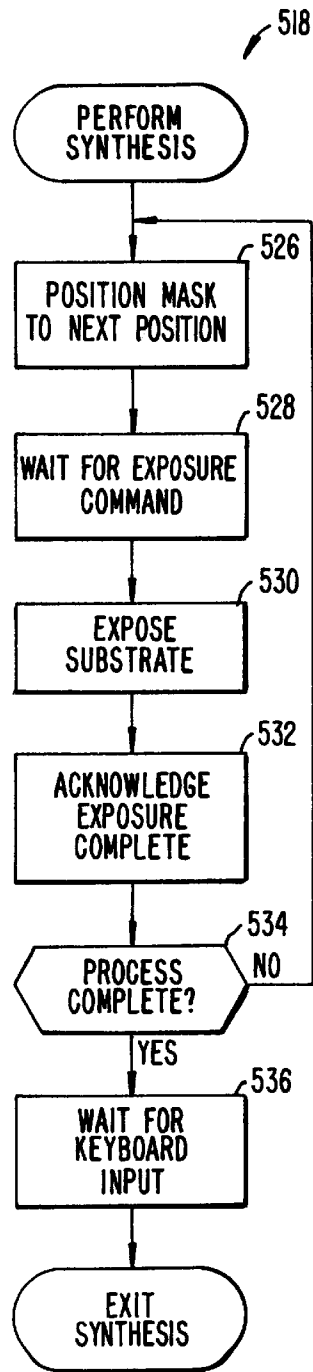


FIG. 24B.

FIG. 25.

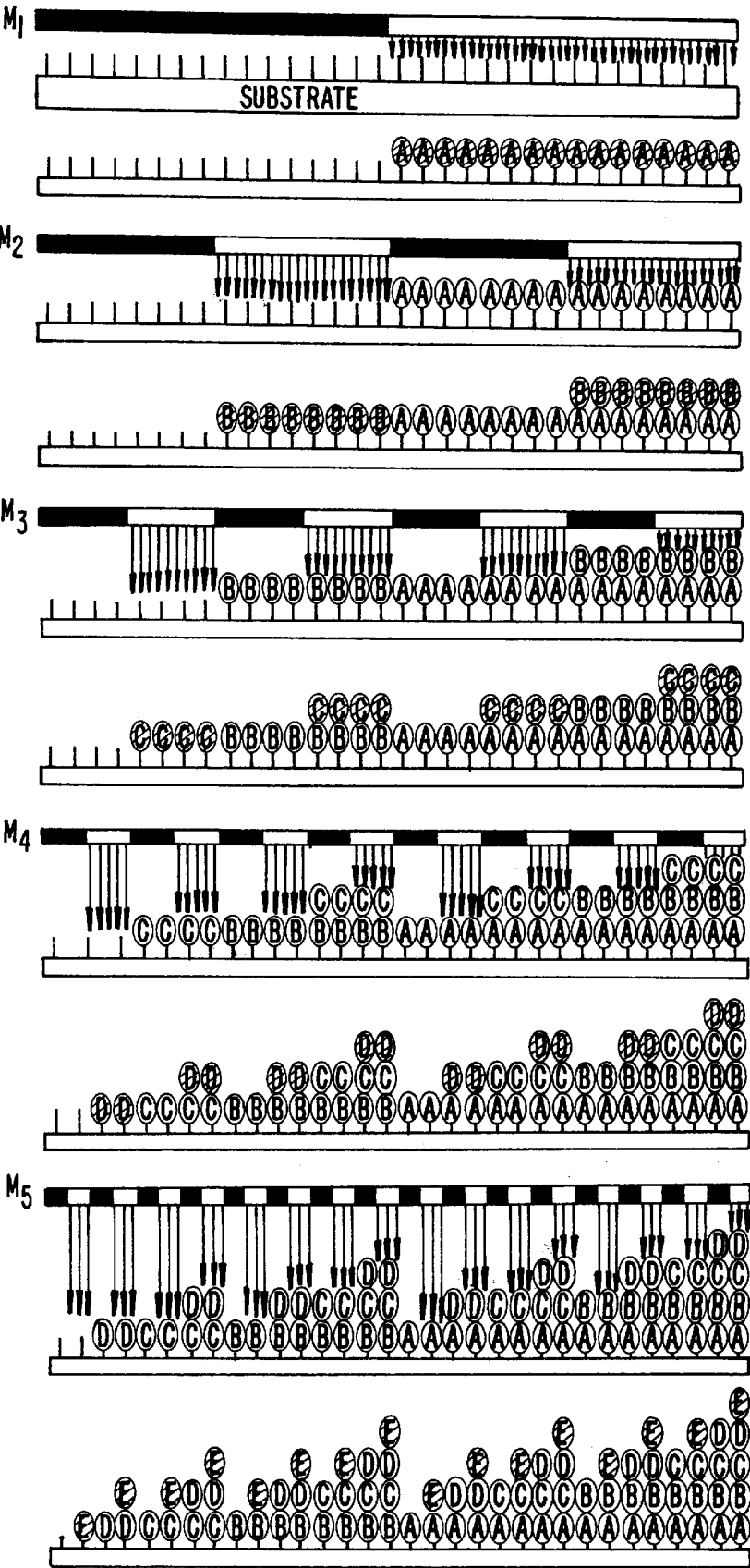


FIG. 26.

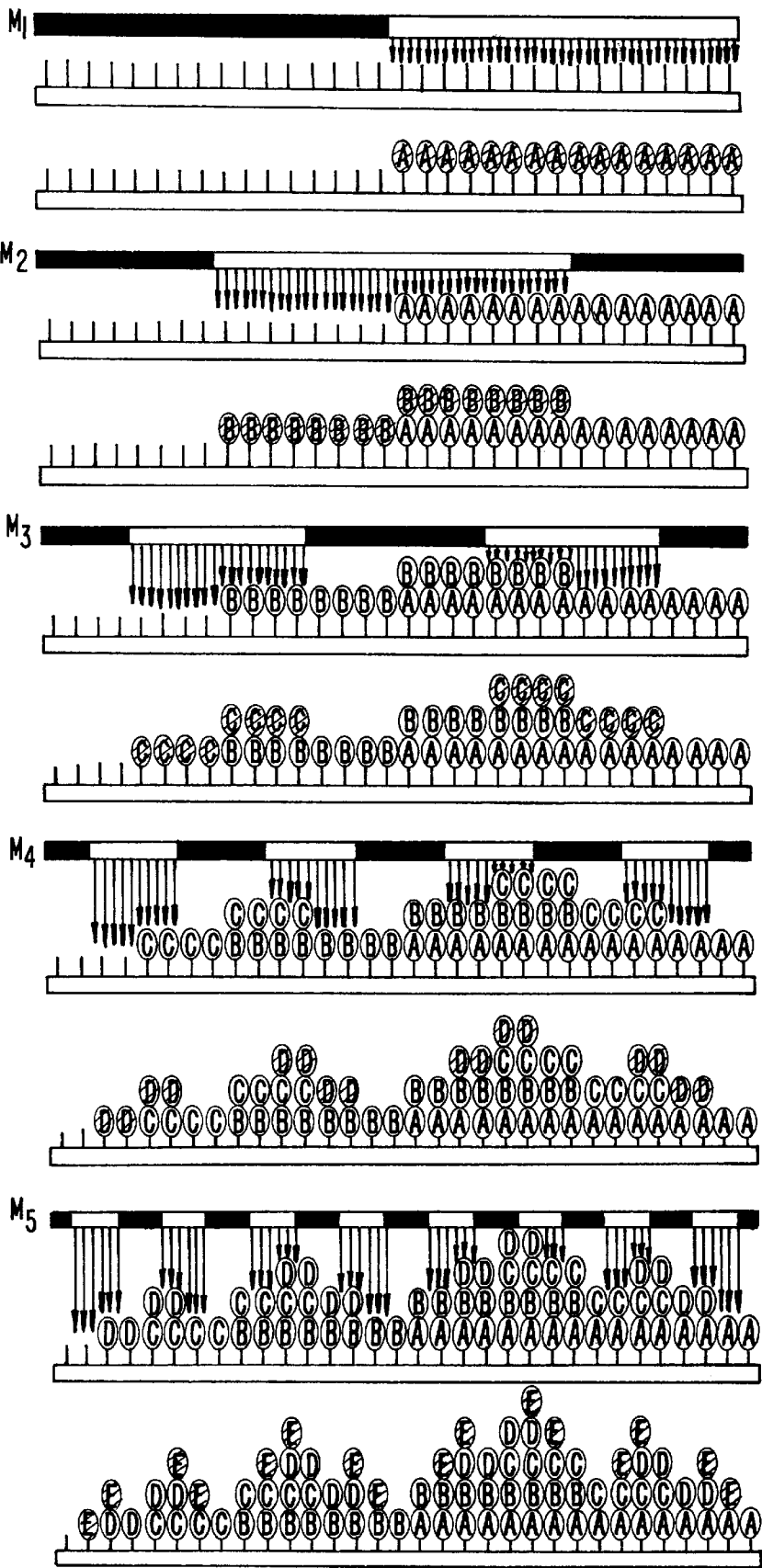
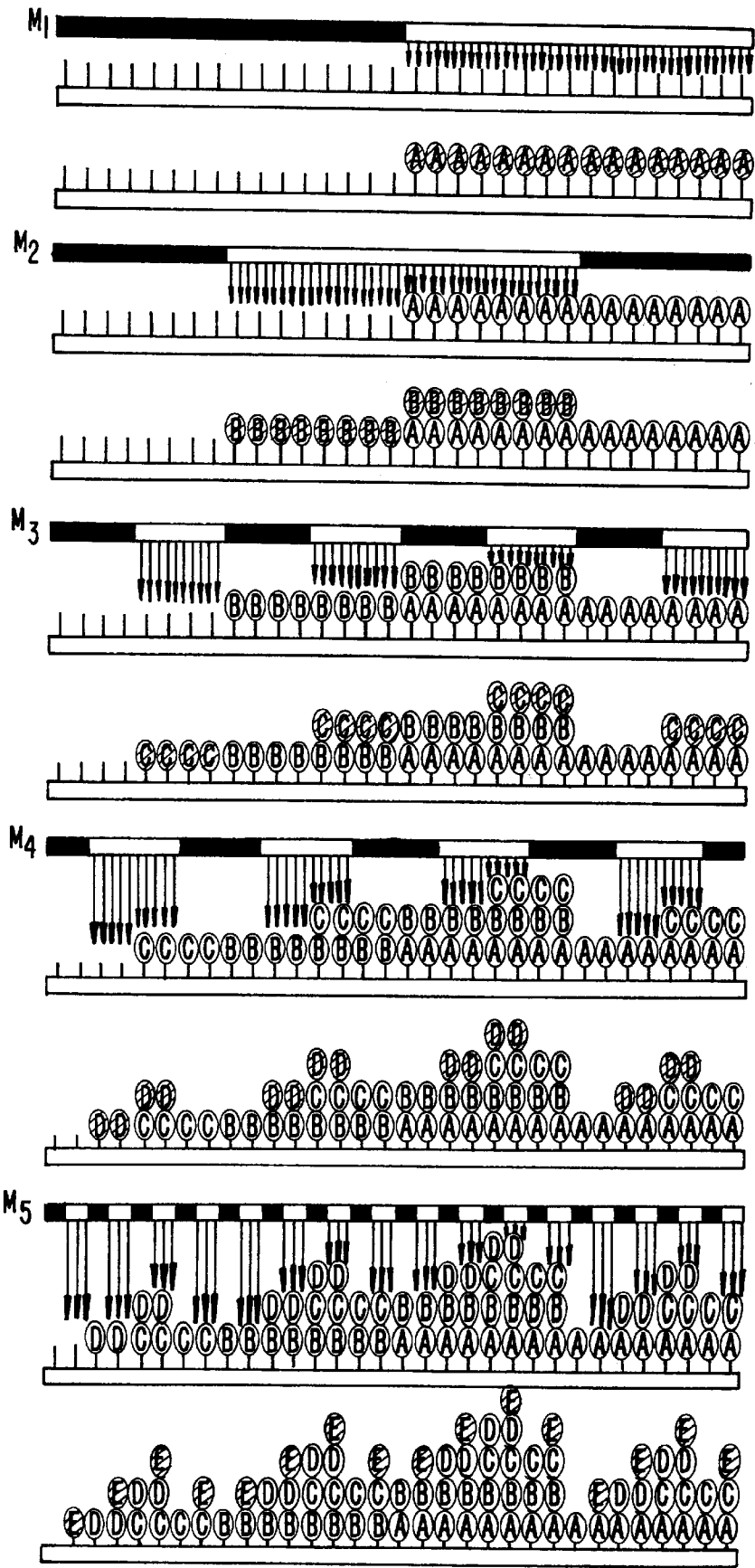


FIG. 27.



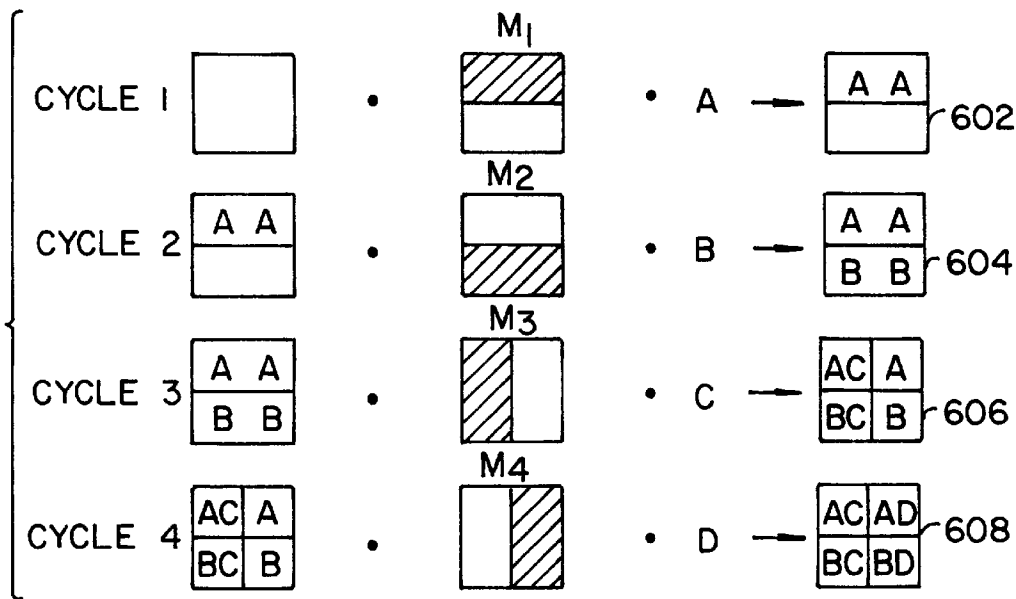


FIG. 28A.

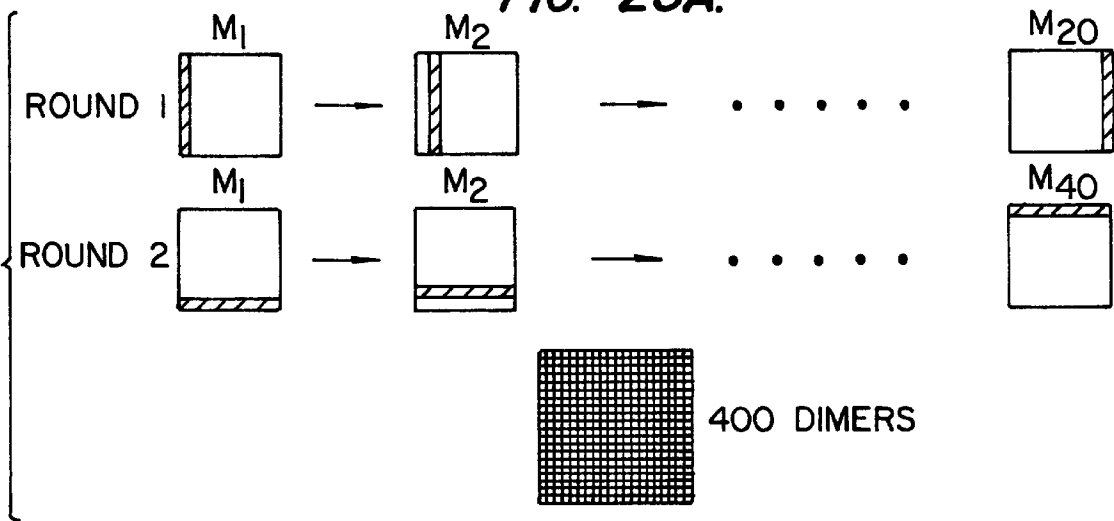


FIG. 28B.

(x, y)

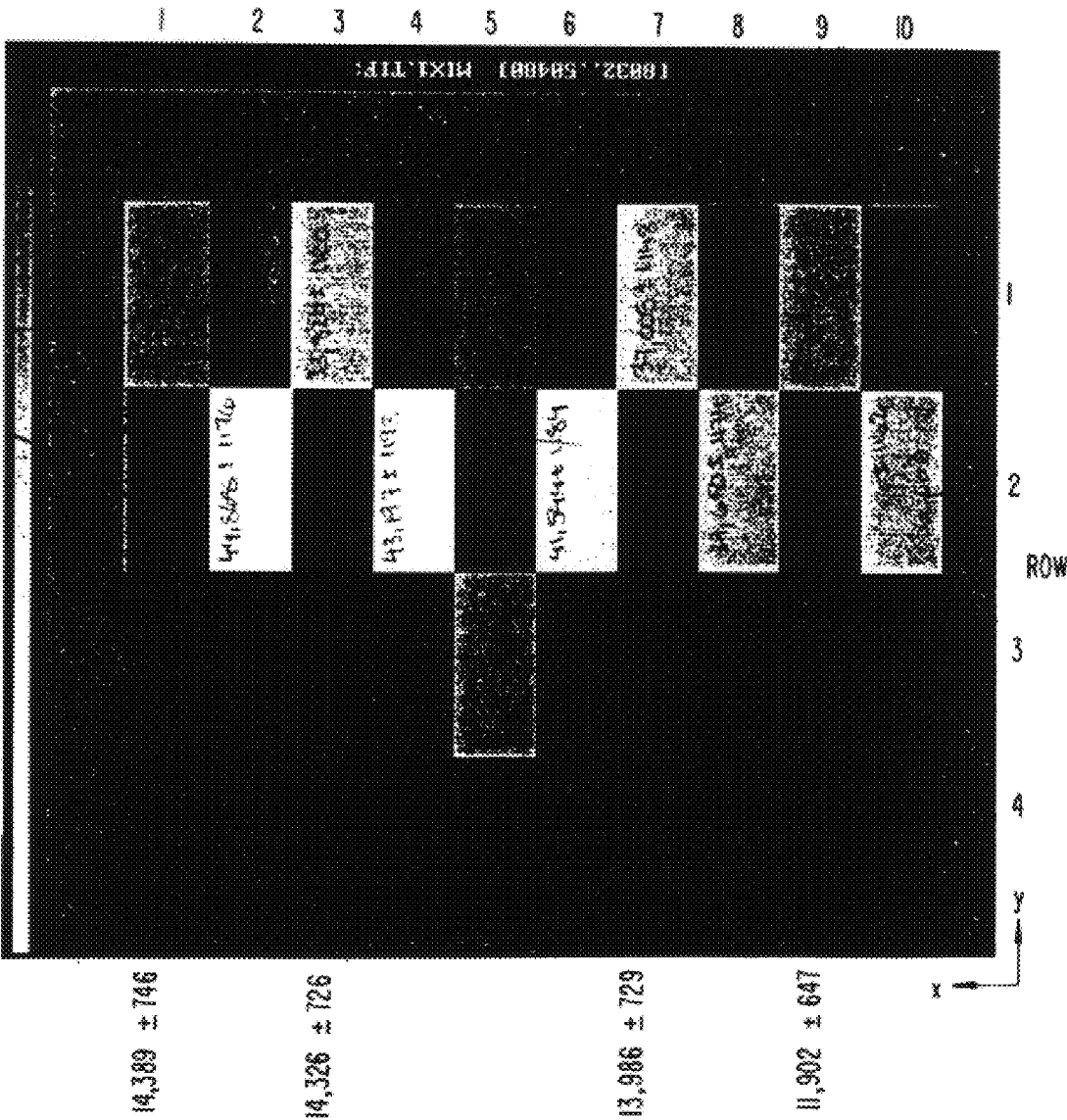










FIG. 30.

STEP	AREA PHOTOLYZED	MASK	COUPLE
1	100%		T
2	100%	"	V
3	100%	"	V
4	100%	"	K
5	50%		F
6 TO 25	Y20	<div><div>20 STEPS</div><div>   </div><div>W NOT OVERLAPPING</div></div>	G, A, R, K, C, M, S D, E, N, Q, F, H W, Y, L, P, V, I, T
26	50%		Q
27	100%		R

WILL GENERATE AN ARRAY OF 4 CLASSES OF PEPTIDES:

- (1) RXKVVT

(2) RQXKVVT

(3) RQXFKVVT

(4) RFXKVVT

WHERE X REPRESENTS SUBSTITUTION OF ALL
20 L-AMINO ACIDS

FIG. 31.

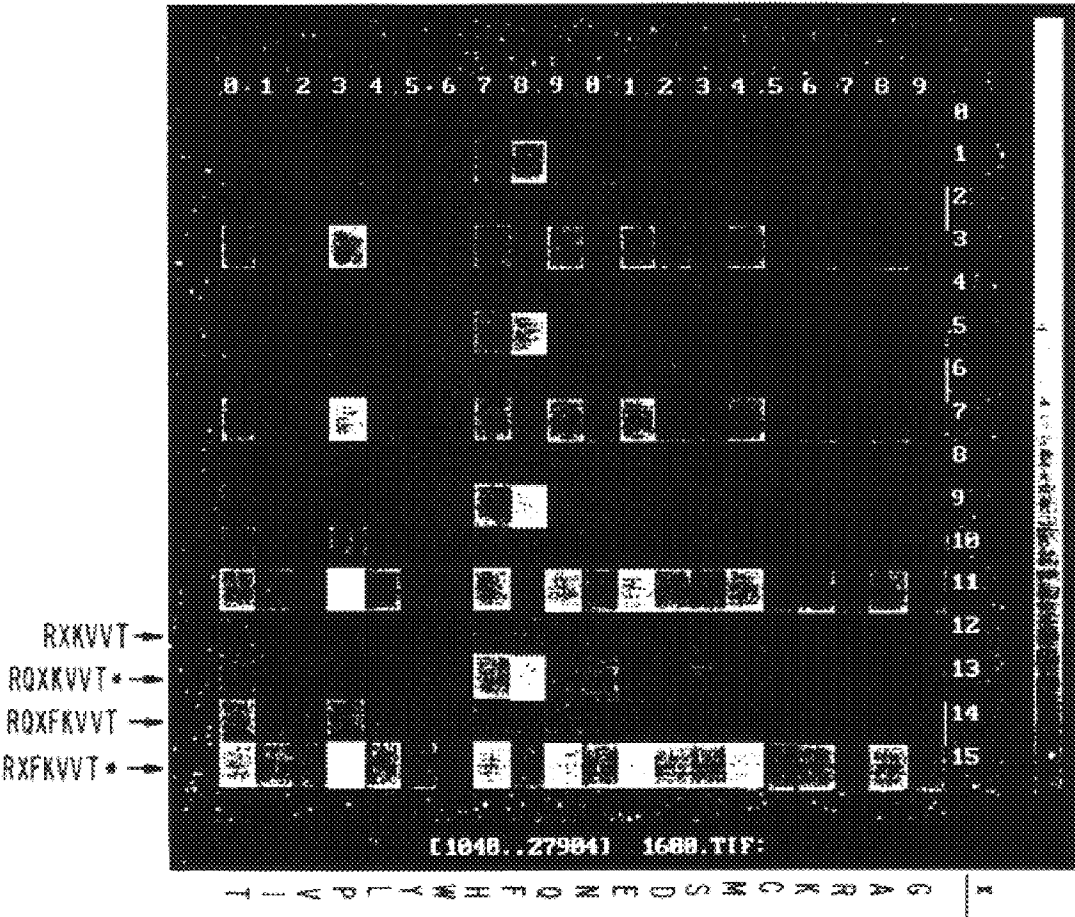


FIG. 32.

U.S. Patent

Jul. 16, 2002

Sheet 34 of 42

US 6,420,169 B1

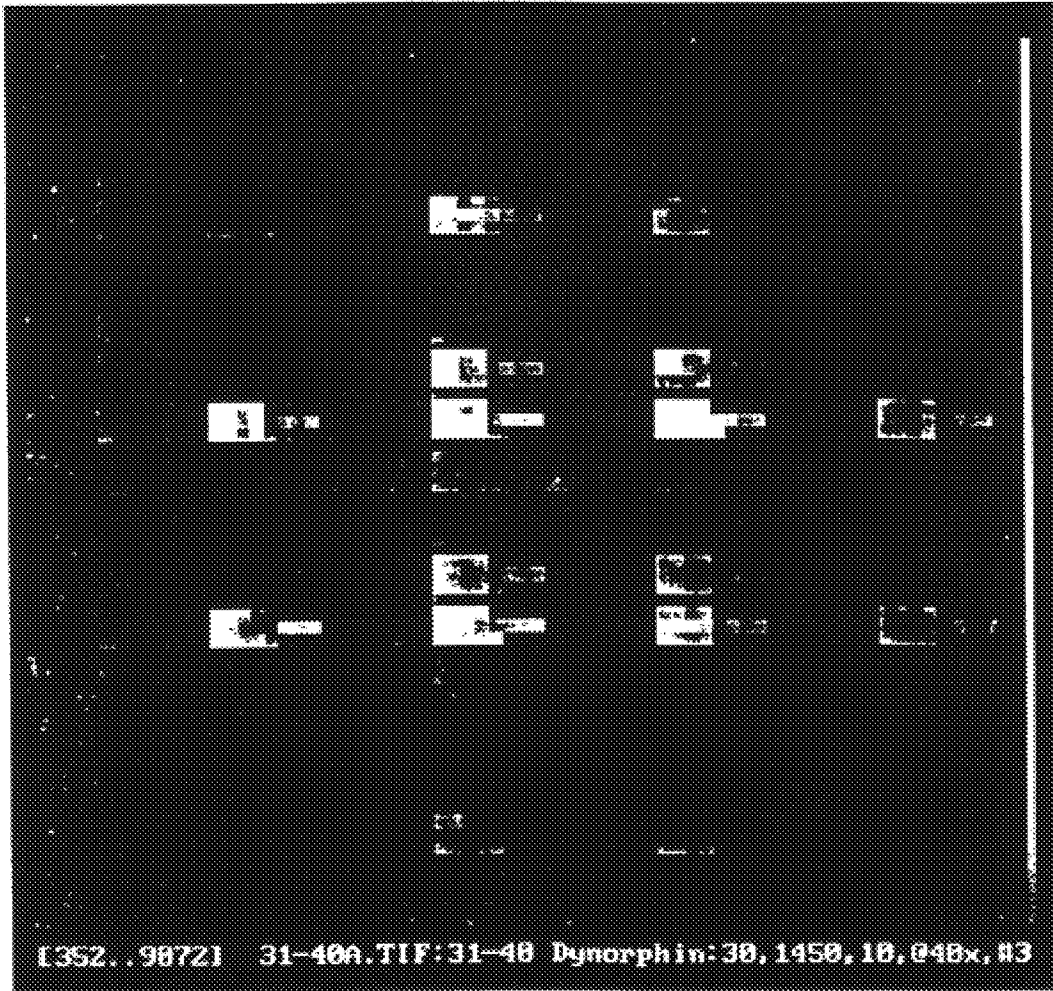


FIG. 33.

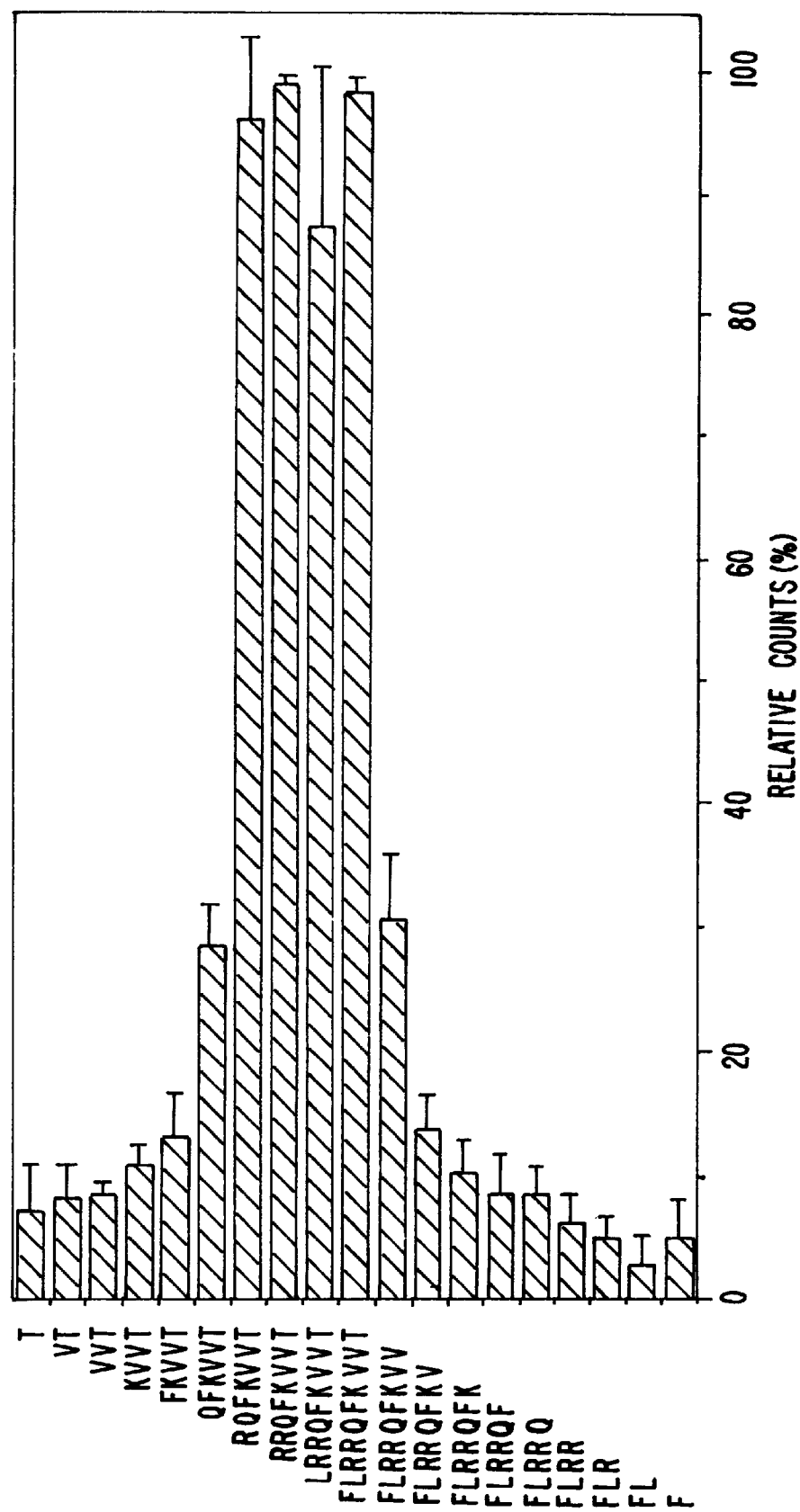


FIG. 34.

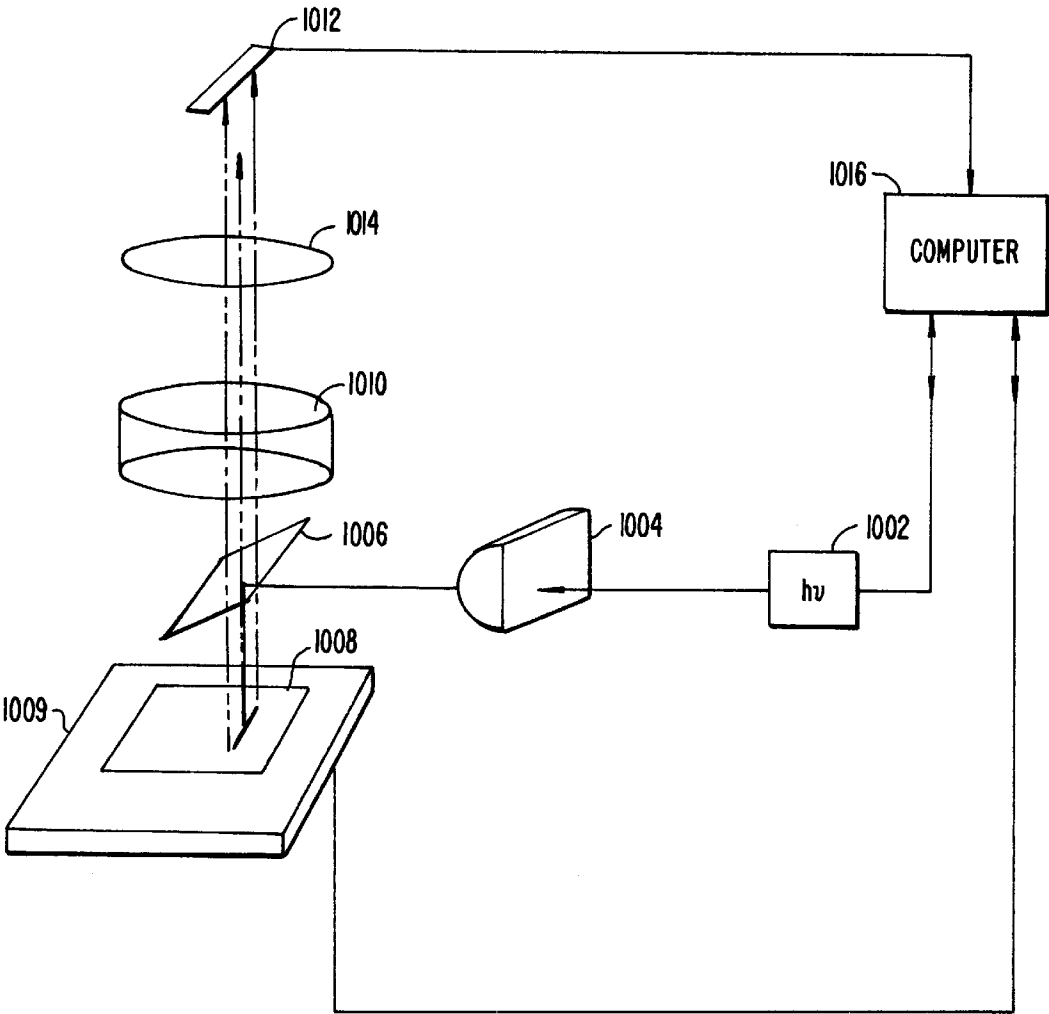


FIG. 35.

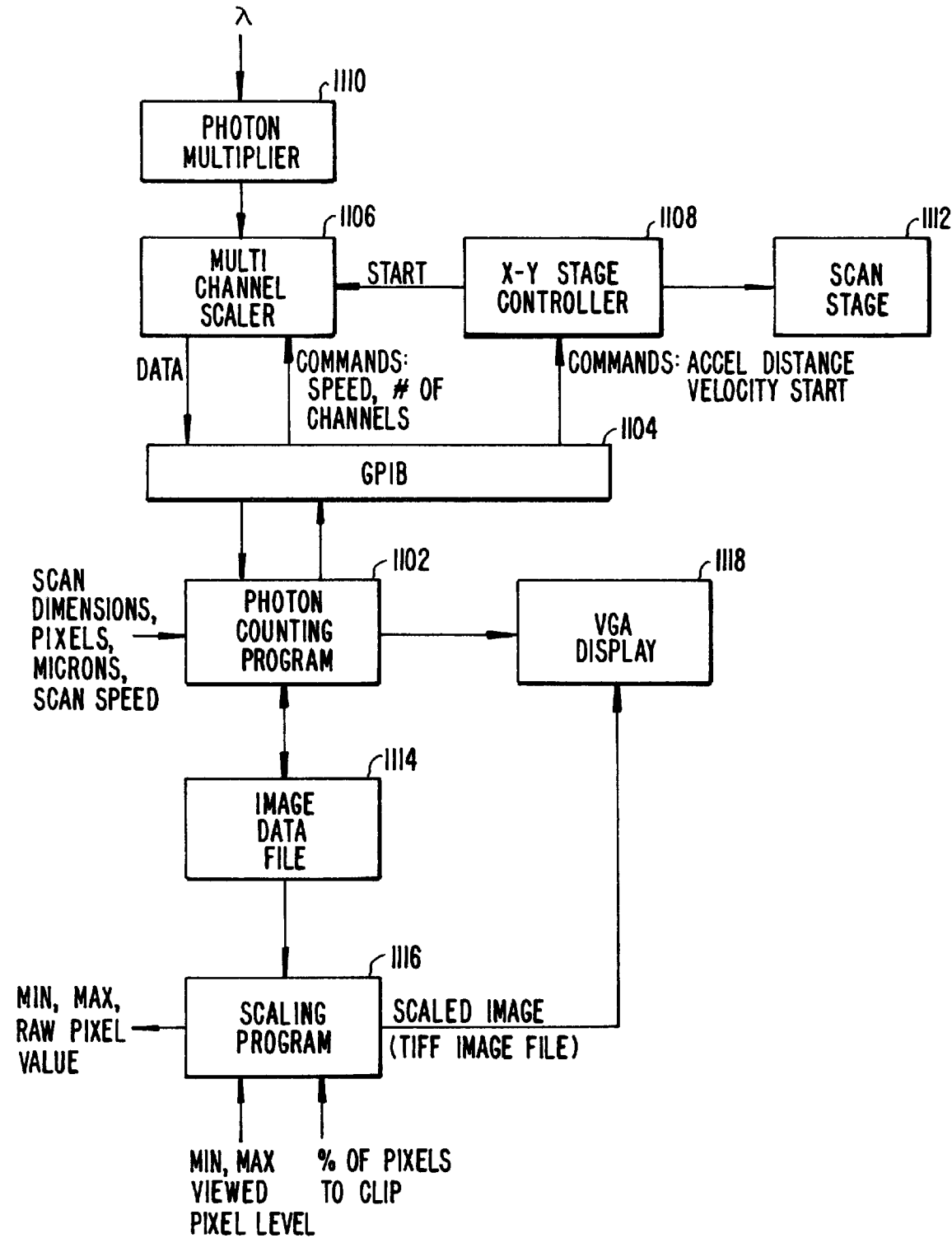


FIG. 36.

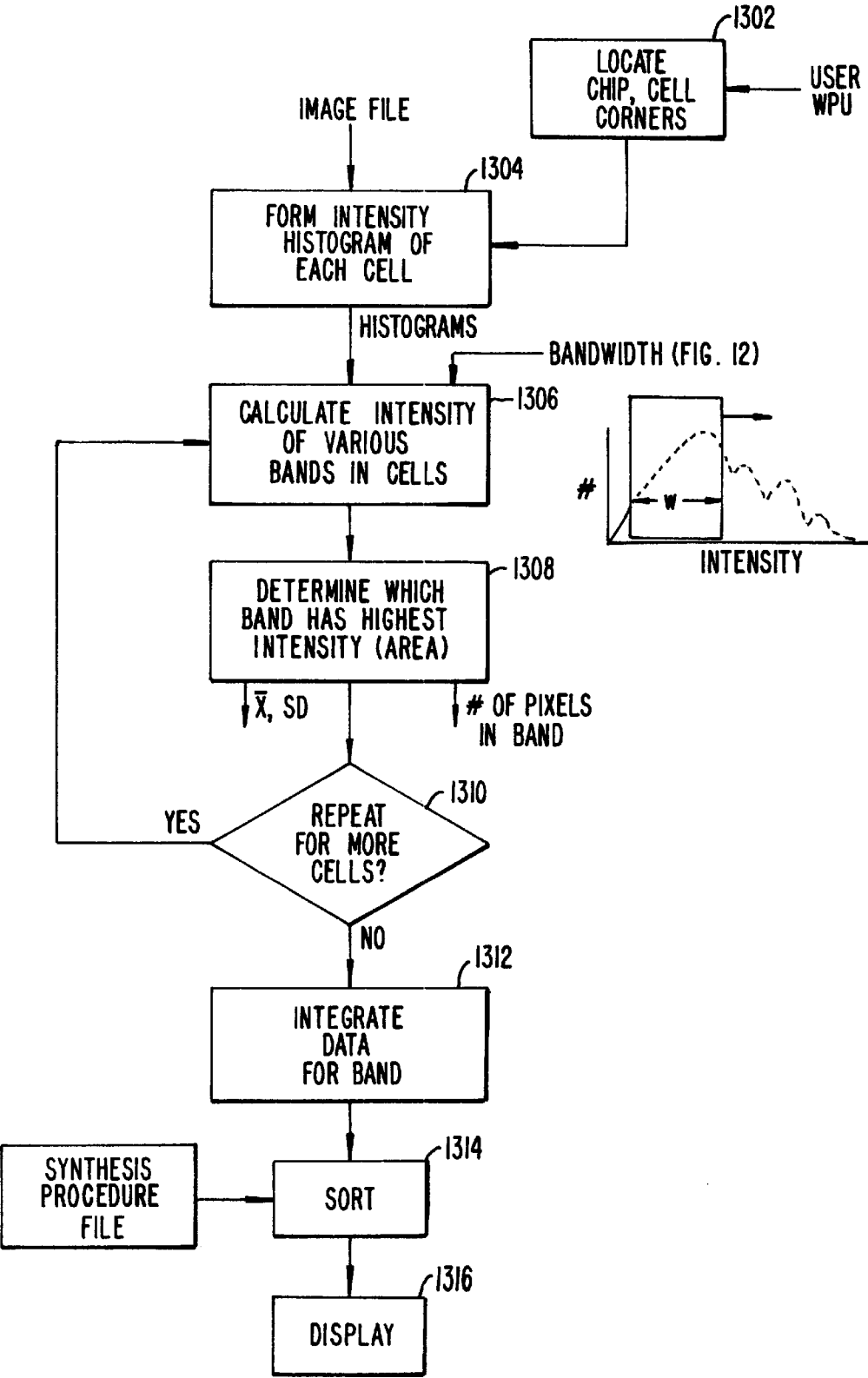


FIG. 37.

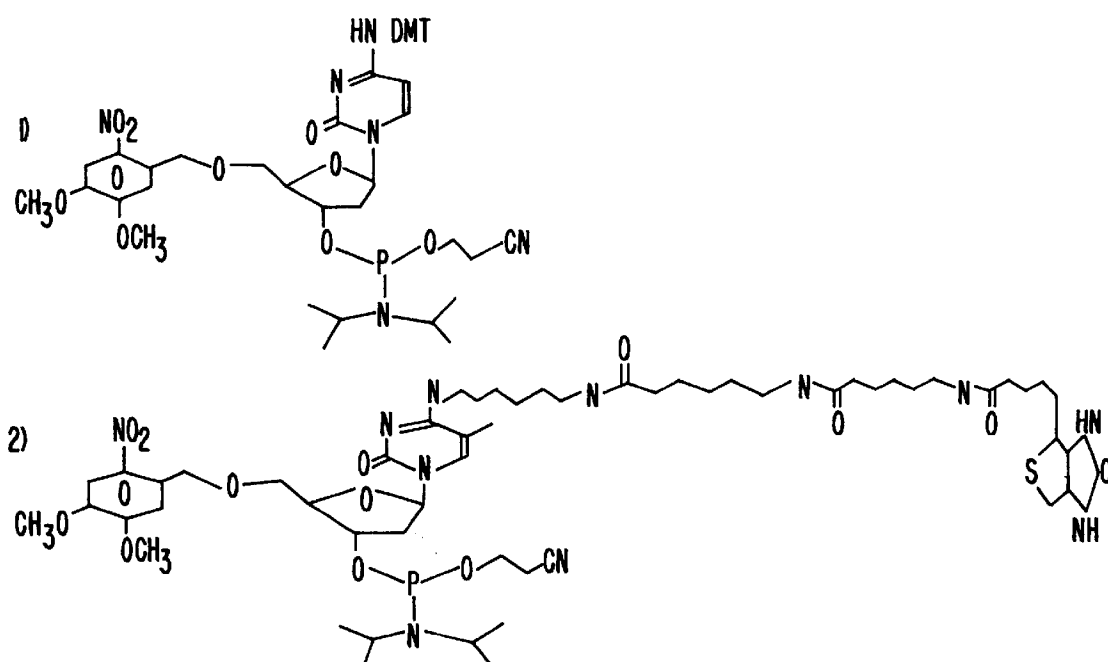
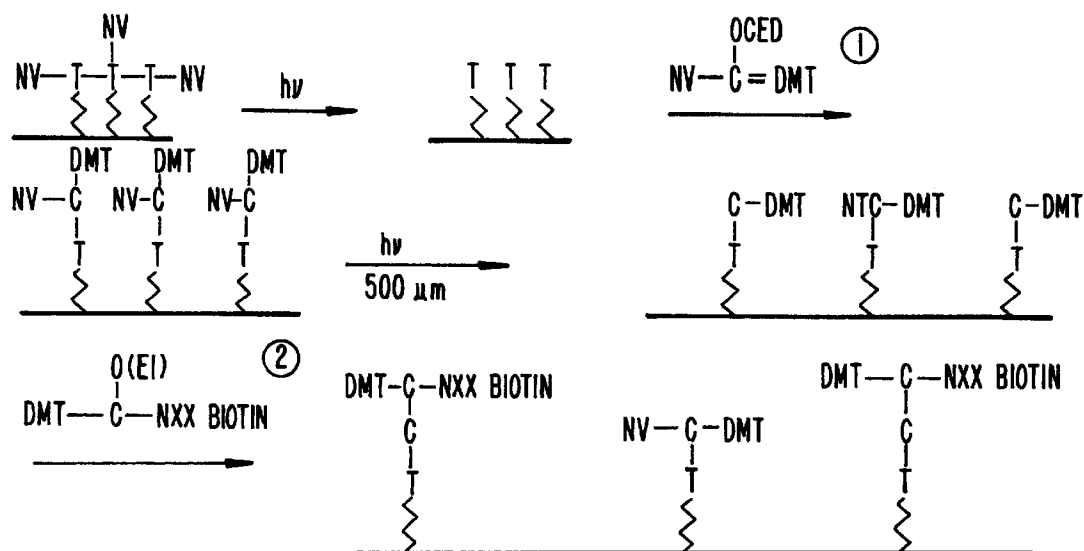
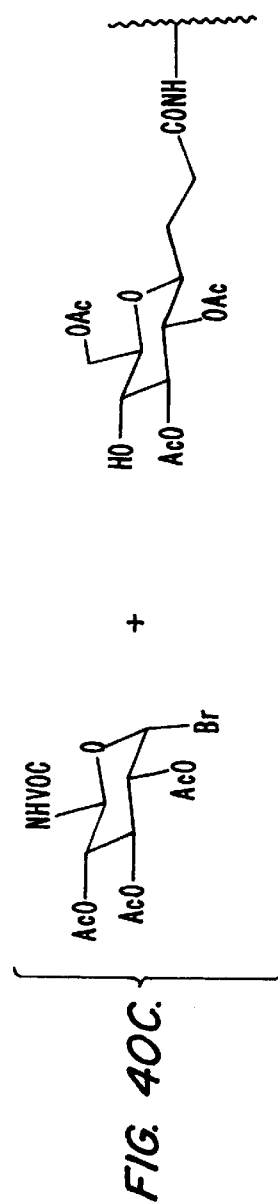
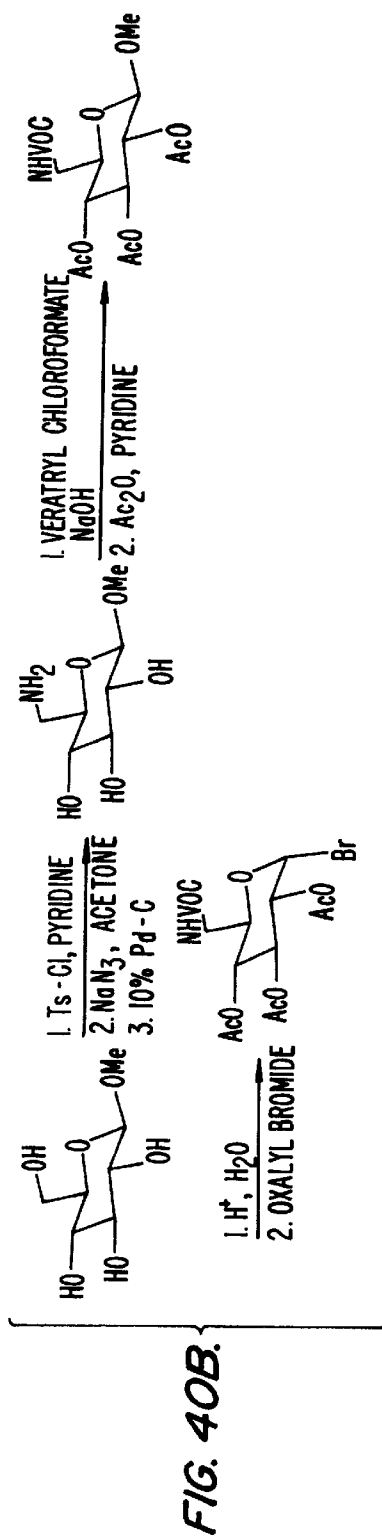
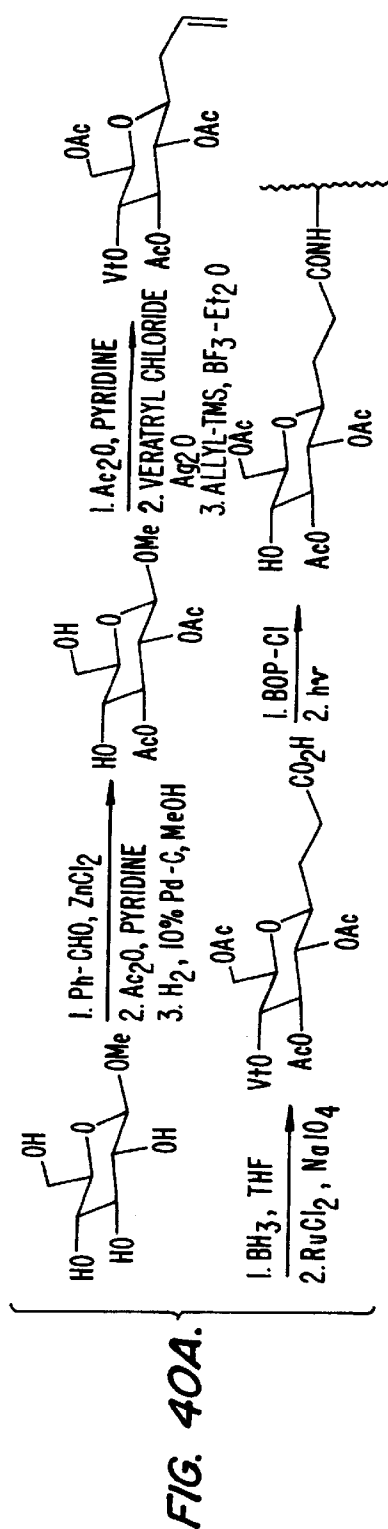
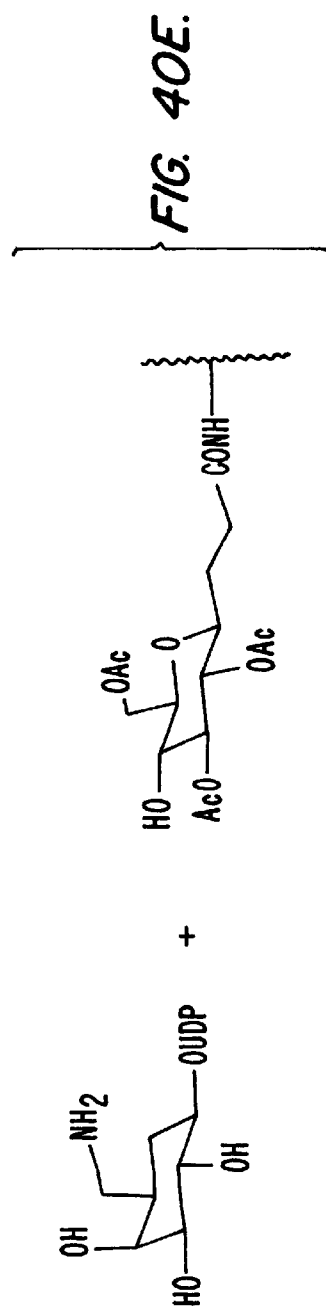
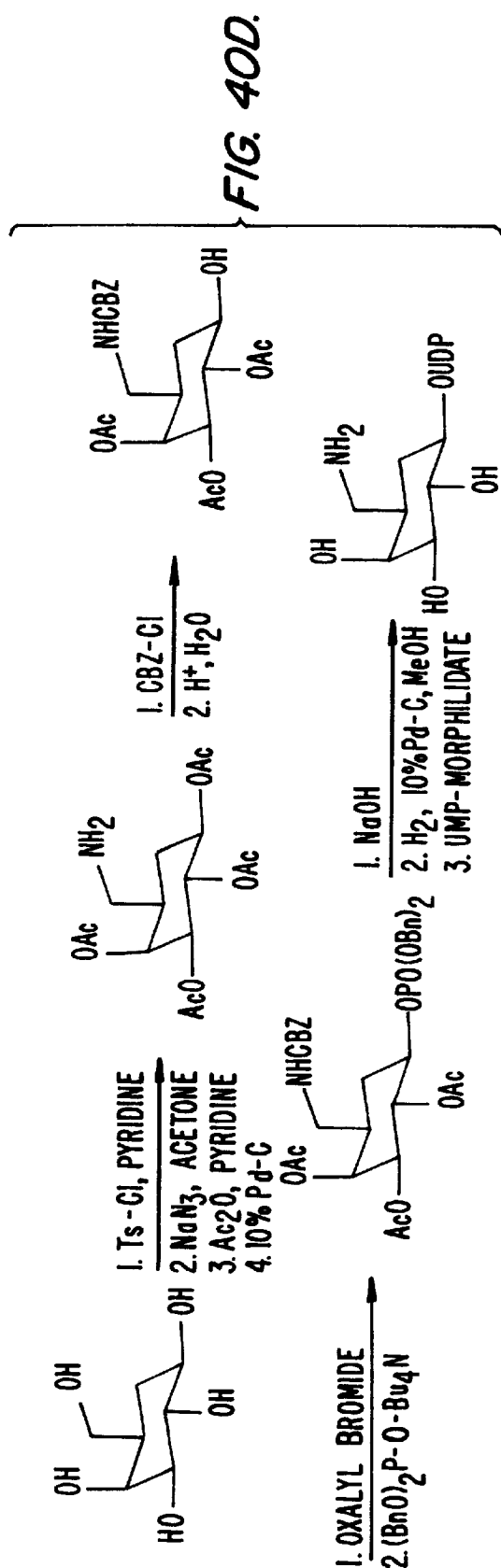


FIG. 38.



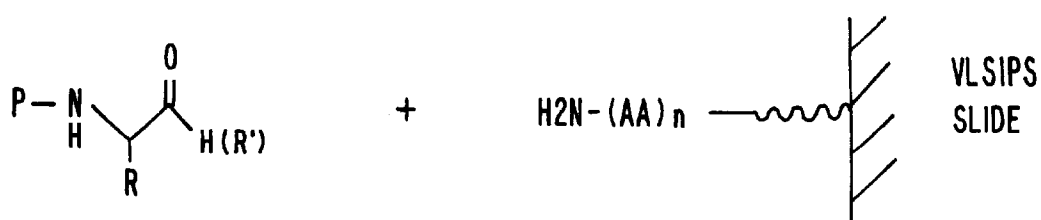


U.S. Patent

Jul. 16, 2002

Sheet 42 of 42

US 6,420,169 B1



WHERE R = AMINO ACID SIDE CHAIN OR OTHER DERIVATIVES

R' = ALKYL

P = PHOTO LABILE PROTECTING GROUP

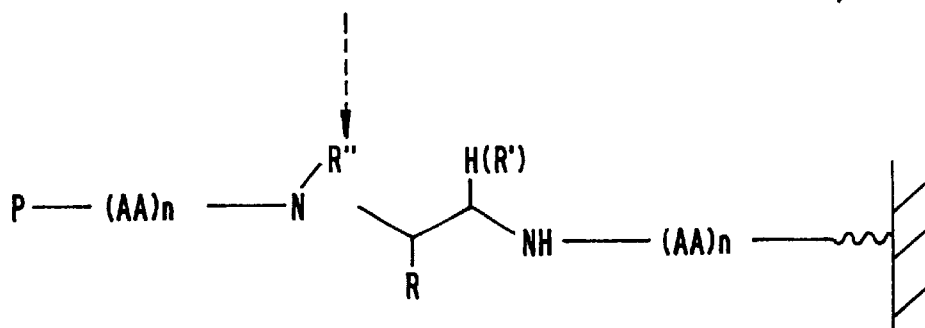
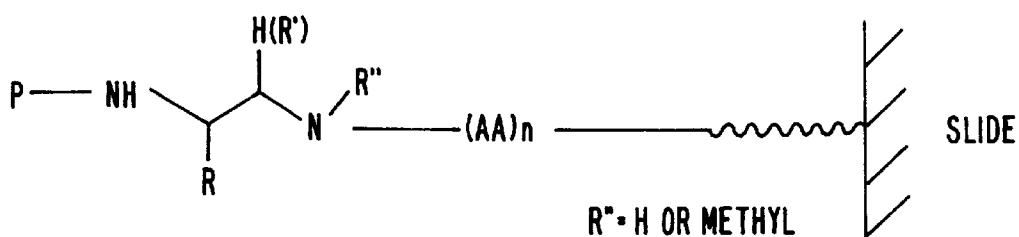
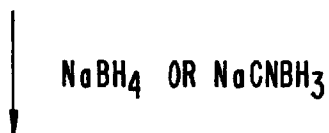
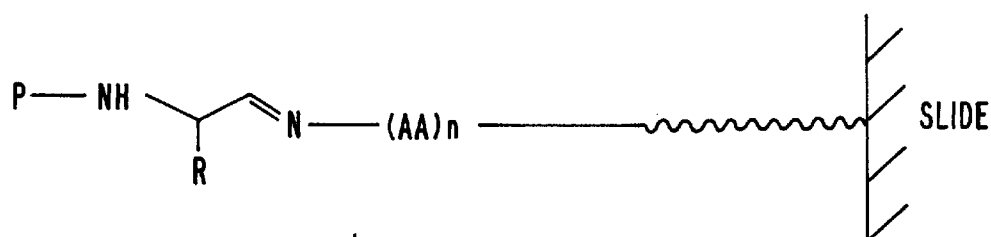
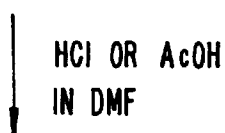


FIG. 41.

US 6,420,169 B1

1

APPARATUS FOR FORMING POLYNUCLEOTIDES OR POLYPEPTIDES

This application is a continuation of Ser. No. 07/805,727, filed Dec. 6, 1991, now U.S. Pat. No. 5,424,186; which is a continuation-in-part of U.S. application Ser. No. 07/492,462, filed Mar. 7, 1990, now U.S. Pat. No. 5,143,854; which is a continuation-in-part of U.S. application Ser. No. 07/362,901, filed Jun. 7, 1989, now abandoned. U.S. application Ser. No. 07/805,727, filed Dec. 6, 1991, now U.S. Pat. No. 5,424,186 is also a continuation-in-part of U.S. Ser. No. 07/624,120, filed Dec. 6, 1990, now abandoned; which is a continuation-in-part of U.S. application Ser. No. 07/492,462, filed Mar. 7, 1990, now U.S. Pat. No. 5,143,854; which is a continuation-in-part of U.S. application Ser. No. 07/362,901, filed Jun. 7, 1989, now abandoned. All of these applications are incorporated herein by reference for all purposes.

This application is also related to the following U.S. Applications, all of which are incorporated herein by reference for all purposes; U.S. Ser. No. 07/626,730, filed Dec. 6, 1990, now U.S. Pat. No. 5,547,839; U.S. application Ser. No. 07/624,114, filed Dec. 6, 1990, now abandoned; U.S. application Ser. No. 07/796,243, filed Nov. 22, 1991, now U.S. Pat. No. 5,384,261; U.S. application Ser. No. 07/796,947, filed Nov. 22, 1991, now U.S. Pat. No. 5,242,974; U.S. application Ser. No. 07/796,727, filed Nov. 22, 1991, now U.S. Pat. No. 5,242,974; and PCT Publication No. WO 90/15070, published Dec. 13, 1990.

COPYRIGHT NOTICE

A portion of the disclosure of this patent document contains material which is subject to copyright protection. The copyright owner has no objection to the facsimile reproduction by anyone of the patent document or the patent disclosure exactly as it appears in the Patent and Trademark Office patent file or records, but otherwise reserves all copyright rights whatsoever.

MICROFICHE APPENDIX

This specification includes microfiche appendices 1, 2 and 3, having 5 sheets with 348 frames.

BACKGROUND OF THE INVENTION

The present invention relates to the field of polymer synthesis. More specifically, the invention provides a reactor system, a masking strategy, photoremovable protecting groups, data collection and processing techniques, and applications for light directed synthesis of diverse polymer sequences on substrates.

Prior methods of preparing large numbers of different polymers have been painstakingly slow when used at a scale sufficient to permit effective rational or random screening. For example, the "Merrifield" method (*J. Am. Chem. Soc.* (1963) 85:2149-2154, which is incorporated herein by reference for all purposes) has been used to synthesize peptides on a solid support. In the Merrifield method, an amino acid is covalently bonded to a support made of an insoluble polymer. Another amino acid with an alpha protecting group is reacted with the covalently bonded amino acid to form a dipeptide. After washing, the protecting group is removed and a third amino acid with an alpha protecting group is added to the dipeptide. This process is continued until a peptide of a desired length and sequence is obtained. Using the Merrifield method, it is not economically practical to synthesize more than a handful of peptide sequences in a day.

2

To synthesize larger numbers of polymer sequences, it has also been proposed to use a series of reaction vessels for polymer synthesis. For example, a tubular reactor system may be used to synthesize a linear polymer on a solid phase support by automated sequential addition of reagents. This method still does not enable the synthesis of a sufficiently large number of polymer sequences for effective economical screening, i.e., for purposes of drug discovery.

Methods of preparing a plurality of polymer sequences are also known in which a container encloses a known quantity of reactive particles, the particles being larger in size than foramina of the container. The containers may be selectively reacted with desired materials to synthesize desired sequences of product molecules. As with other methods known in the art, this method cannot practically be used to synthesize a sufficient variety of polypeptides for effective screening.

Other techniques have also been described. These methods include the synthesis of peptides on 96 plastic pins which fit the format of standard microtiter plates. Unfortunately, while these techniques have been somewhat useful, substantial problems remain. For example, these methods continue to be limited in the diversity of sequences which can be economically synthesized and screened.

From the above, it is seen that an improved method and apparatus for synthesizing a variety of chemical sequences at known locations is desired.

SUMMARY OF THE INVENTION

Methods, apparatus, and compositions for synthesis and use of diverse polymer sequences on a substrate are disclosed, as well as applications thereof.

In one preferred embodiment, linker molecules are provided on a substrate. A terminal end of the linker molecules is provided with a reactive functional group protected with a photoremovable protective group. Using lithographic methods, the photoremovable protecting group is exposed to light and removed from the linker molecules in the first selected regions. The substrate is then washed or otherwise contacted with a first monomer that reacts with exposed functional groups on the linker molecules. In a preferred embodiment, the monomer is an amino acid containing a photoremovable protecting group at its amino or carboxy terminus and the linker molecule terminates in an amino or carboxy acid group bearing a photoremovable protective group.

A second set of selected regions is, thereafter, exposed to light and the photoremovable protecting group on the linker molecule/protected amino acid is removed at the second set of regions. The substrate is then contacted with a second monomer containing a photoremovable protecting group for reaction with exposed functional groups. This process is repeated to selectively apply monomers until polymers of a desired length and desired chemical sequence are obtained. Photolabile groups are then optionally removed and the sequence is, thereafter, optionally capped. Side chain protective groups, if present, are also removed.

By using the lithographic techniques disclosed herein, it is possible to direct light to relatively small and precisely known locations on the substrate. It is, therefore, possible to synthesize polymers of a known chemical sequence at known locations on the substrate. The general version of this technique is termed Very Large Scale Immobilized Polymer Synthesis (VLSIPS™).

The resulting substrate will have a variety of uses including, for example, screening large numbers of poly-

US 6,420,169 B1

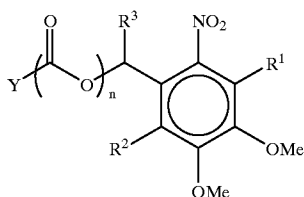
3

mers for biological activity. To screen for biological activity, the substrate is exposed to one or more receptors such as an antibody, whole cells, receptors on vesicles, lipids, or any one of a variety of other receptors. The receptors are preferably labeled with, for example, a fluorescent marker, radioactive marker, or a labeled antibody reactive with the receptor. The location of the marker on the substrate is detected with, for example, photon detection or autoradiographic techniques. Through knowledge of the sequence of the material at the location where binding is detected, it is possible to quickly determine which sequence binds with the receptor and, therefore, the technique can be used to screen large numbers of peptides. Other possible applications of the inventions herein include diagnostics in which various antibodies for particular receptors would be placed on a substrate and, for example, blood sera would be screened for immune deficiencies. Still further applications include, for example, selective "doping" of organic materials in semiconductor devices, i.e., the introduction of selected impurities into the device, and the like.

According to one aspect of the invention, an improved reactor system for synthesis of diverse polymer sequences on a substrate is provided. According to this embodiment the invention provides for a reactor for contacting reaction fluids to a substrate; a system for delivering selected reaction fluids to the reactor; a translation stage for moving a mask or substrate from at least a first relative location relative to a second relative location; a light for illuminating the substrate through a mask at selected times; and an appropriately programmed digital computer for selectively directing a flow of fluids from the reactor system, selectively activating the translation stage, and selectively illuminating the substrate so as to form a plurality of diverse polymer sequences on the substrate at predetermined locations.

The invention also provides a technique for selection of linker molecules using the VLSIPSTM synthesis technique. According to this aspect of the invention, the invention provides a method of screening a plurality of linker polymers for use in binding affinity studies. The invention includes the steps of forming a plurality of linker polymers on a substrate in selected regions, the linker polymers are formed by the steps of recursively: (1) on a surface of a substrate, irradiating a portion of the selected regions to remove a protecting group, and contacting the surface with a monomer; (2) contacting the plurality of linker polymers with a ligand; and (3) contacting the ligand with a labeled receptor.

According to another aspect of the invention, improved photoremovable protecting groups are provided. According to this aspect of the invention a compound having the formula:



wherein $n=0$ or 1 ; Y is selected from the group consisting of an oxygen of the carboxyl group of a natural or unnatural amino acid, an amino group of a natural or unnatural amino acid, or the C-5' oxygen group of a natural or unnatural deoxyribonucleic or ribonucleic acid; R^1 and R^2 independently are a hydrogen atom, a lower alkyl, aryl, benzyl,

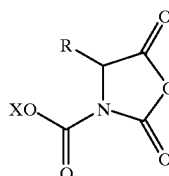
4

halogen, hydroxyl, alkoxy, thiol, thioether, amino, nitro, carboxyl, formate, formamido, sulfido, or phosphido group; and R^3 is a alkoxy, alkyl, aryl, hydrogen, or alkenyl group is provided.

The invention also provides improved masking techniques for VLSIPS. According to one aspect of the masking technique, the invention provides an ordered method for forming a plurality of polymer sequences by sequential addition of reagents comprising the step of serially protecting and deprotecting portions of the plurality of polymer sequences for addition of other portions of the polymer sequences using a combinatorial synthesis strategy.

Improved data collection equipment and techniques are also provided. According to one embodiment, the instrumentation provides a system for determining affinity of a receptor to a ligand comprising: means for applying light to a surface of a substrate, the substrate comprising a plurality of ligands at predetermined locations, the means for applying light providing simultaneous illumination at a plurality of the predetermined locations; and an array of detectors for detecting fluorescence at the plurality of predetermined locations. The invention further provides for improved data analysis techniques including the steps of exposing fluorescently labelled receptors to a substrate, the substrate comprising a plurality of ligands in regions at known locations; at a plurality of data collection points within each of the regions, determining an amount of fluorescence from the data collection points; removing the data collection points deviating from a predetermined statistical distribution; and determining a relative binding affinity of the receptor from remaining data collection points.

Protected amino acid N-carboxy anhydrides for use in polymer synthesis are also disclosed. According to this aspect of the invention, a compound having the following formula is provided:



where R is a side chain of a natural or unnatural amino acid and X is a photoremovable protecting group.

A further understanding of the nature and advantages of the inventions herein may be realized by reference to the remaining portions of the specification and the attached drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGS. 1 to 7 illustrate masking, irradiation, and coupling of monomers;

FIGS. 8A and 8B are fluorescence curves for NVOC slides not exposed and exposed to light respectively;

FIGS. 9A–9D are fluorescence plots of slides exposed through 100 μm , 50 μm , 20 μm and 10 μm masks;

FIG. 10 illustrates fluorescence of a slide with the peptide YGGFL (SEQ. ID NO:1) on selected region on its surface which has been exposed to labeled Herz antibodies specific for the sequence;

FIG. 11 schematically illustrates one example of light-directed peptide synthesis;

FIG. 12 is a fluorescence plot of YGGFL (SEQ. ID NO:1) and PGGFL synthesized in a 50 μm checkerboard pattern;

US 6,420,169 B1

5

FIGS. 13A–13D illustrate formation and screening of a checkerboard pattern of YGGFL (SEQ. ID NO:1) and GGFL (SEQ. ID NO:15);

FIG. 14 is a fluorescence plot of YPGGFL (SEQ. ID NO:3) and YGGFL (SEQ. ID NO:1) synthesized in a 50 μ m checkerboard pattern;

FIGS. 15A and 15B illustrate the mapping of 16 sequences synthesized on two different glass slides;

FIG. 16 is a fluorescence plot of the slide illustrated in FIG. 15A;

FIG. 17 is a fluorescence plot of the slide illustrated in FIG. 15B;

FIG. 18 is a fluorescence plot of an experiment which produced 4,096 compounds;

FIG. 19 is a fluorescence plot of a substrate on which 65,536 different compounds were formed;

FIGS. 20A and 20B show a tripeptide used in a fluorescence energy-transfer substrate assay and that substrate after cleavage;

FIGS. 21A and 21B are fluorescence plots generated with fluorescence energy-transfer substrate assays;

FIGS. 22A and 22B illustrate alternative embodiments of a reactor system for forming a plurality of polymers on a substrate;

FIG. 23 schematically illustrates an automated system for synthesizing diverse polymer sequences;

FIGS. 24A and 24B illustrate operation of a program for polymer synthesis;

FIG. 25 is a schematic illustration of a “pure” binary masking strategy;

FIG. 26 is a schematic illustration of a gray code binary masking strategy;

FIG. 27 is a schematic illustration of a modified gray code binary masking strategy;

FIG. 28A schematically illustrates a masking strategy for a four step synthesis;

FIG. 28B schematically illustrates synthesis of 400 peptide dimers of genetically coded amino acids;

FIG. 29 is a coordinate map for the ten-step binary synthesis;

FIG. 30 is a fluorescence plot of a 4 \times 10 array of peptides having sequences similar to dynorphin B;

FIG. 31 illustrates a strategy for producing an array of peptides related to the dynorphin B sequence;

FIG. 32 is a fluorescence plot of an array of peptides produced according to the strategy illustrated in FIG. 31;

FIG. 33 is a fluorescence plot of an array of peptides containing various deletions from the dynorphin B sequence;

FIG. 34 is a plot of the relative binding affinities of an anti-dynorphin B monoclonal antibody to various sequences produced on the substrate shown in FIG. 33;

FIG. 35 schematically illustrates a data collection system;

FIG. 36 is a block diagram illustrating the architecture of the data collection system;

FIG. 37 is a flow chart illustrating operation of software for the data collection/analysis system;

FIG. 38 schematically illustrates one example of light-directed oligonucleotide synthesis;

FIGS. 39A–39C are fluorescence plots demonstrating hybridization, dehybridization and rehybridization between immobilized poly A and poly T;

6

FIGS. 40A–40E illustrate a synthesis strategy for forming polysaccharides in accordance with the present invention; and

FIG. 41 illustrates the introduction of a 1reduced amide bond into a growing peptide.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

CONTENTS

I. Definitions

II. General

A. Deprotection and Addition

1. Example—Polymer Synthesis

2. Example

3. Example—Slide Preparation

4. Example—Synthesis of a Dimer of an Amino-propyl Group and a Fluorescent Group

5. Example—Removal of NVOC and Attachment of a Marker

6. Example—Use of a Mask in Removal of NVOC

7. Example

8. Example

B. Antibody recognition

1. Example—Attachment of YGGFL and Subsequent Exposure to Herz Antibody and Goat Antimouse

2. Example

3. Example—Monomer-by-Monomer Formation of YGGFL and Subsequent Exposure to Labeled Antibody

4. Example—Monomer-by-Monomer Synthesis of YGGFL and YPGGFL

5. Example—Synthesis of an Array of Sixteen Different Amino Acid Sequences and Estimation of Relative Binding Affinity to Herz Antibody

6. Example

7. Example

C. Fluorescence Energy-Transfer Substrate Assays

III. Synthesis

A. Reactor System

B. Binary Synthesis Strategy

1. Example

2. Example

3. Example

4. Example

5. Example

6. Example

7. Example

8. Example

9. Example

10. Example

C. Linker Selection

D. Protecting Groups

1. Use of Photoremovable Protecting Groups During Solid-Phase Synthesis of Peptides

2. Use of Photoremovable Protecting Groups During Solid-Phase Synthesis of Oligonucleotides

E. Amino Acid N-Carboxy Anhydrides Protected with a Photoremovable Group

IV. Data Collection

A. Data Collection System

B. Data Analysis

C. Alternative Embodiments

US 6,420,169 B1

7

V. Other Representative Applications

A. Oligonucleotide Synthesis

1. Example
2. Example
3. Example

B. Oligosaccharide Synthesis

1. Example

C. Caged Binding Member System

D. Fingerprinting for Quality Control

E. β -Amino Acid and D-Amino Acid Monomers

F. Reduced Amide Bonds

VI. Conclusion

I. Definitions

Certain terms used herein are intended to have the following general definitions:

1. Complementary

Refers to the topological compatibility or matching together of interacting surfaces of a ligand molecule and its receptor.

Thus, the receptor and its ligand can be described as complementary, and furthermore, the contact surface characteristics are complementary to each other.

2. Epitope

The portion of an antigen molecule which is delineated by the area of interaction with the subclass of receptors known as antibodies.

3. Ligand

A ligand is a molecule that is recognized by a particular receptor. Examples of ligands that can be investigated by this invention include, but are not restricted to, agonists and antagonists for cell membrane receptors, toxins and venoms, viral epitopes, hormones (e.g., steroids, etc.), hormone receptors, opiates, peptides, enzymes, enzyme substrates, cofactors, drugs, lectins, sugars, oligonucleotides, nucleic acids, oligosaccharides, proteins, and monoclonal antibodies.

4. Monomer

A member of the set of small molecules which can be joined together to form a polymer. The set of monomers includes but is not restricted to, for example, the set of common L-amino acids, the set of D-amino acids, the set of natural or synthetic amino acids, the set of nucleotides and the set of pentoses and hexoses. As used herein, monomer refers to any member of a basis set for synthesis of a polymer. For example, dimers of the 20 naturally occurring L-amino acids form a basis set of 400 monomers for synthesis of polypeptides. Different basis sets of monomers may be used in any of the successive steps in the synthesis of a polymer. Furthermore, each of the sets may include protected members which are modified after synthesis.

5. Peptide

A polymer in which the monomers are alpha amino acids and which are joined together through amide bonds, alternatively referred to as a polypeptide. In the context of this specification it should be appreciated that the amino acids may, for example, the L-optical isomer or the D-optical isomer. Peptides are often two or more amino acid monomers long, and often 4 or more amino acids long, often 5 or more amino acids long, often 10 or more amino acids long, often 15 or more amino acids long, and often 20 or more amino acid monomers long, for example. Standard abbreviations for amino acids are used (e.g., P for proline). These abbreviations are included in Stryer, *Biochemistry*, Third Ed., 1988, which is incorporated herein by reference for all purposes.

8

6. Radiation

Energy which may be selectively applied including energy having a wavelength of between 10^{-14} and 10^4 meters including, for example, electron beam radiation, gamma radiation, x-ray radiation, ultra-violet radiation, visible light, infrared radiation, microwave radiation, and radio waves. "Irradiation" refers to the application of radiation to a surface.

7. Receptor

A molecule that has an affinity for a given ligand. Receptors may be naturally-occurring or synthetic molecules. Also, they can be employed in their unaltered state or as aggregates with other species. Receptors may be attached, covalently or noncovalently, to a binding member, either directly or via a specific binding substance. Examples of receptors which can be employed by this invention include, but are not restricted to, antibodies, cell membrane receptors, monoclonal antibodies and antisera reactive with specific antigenic determinants (such as on viruses, cells, or other materials), drugs, polynucleotides, nucleic acids, peptides, cofactors, lectins, sugars, polysaccharides, cells, cellular membranes, and organelles. Receptors are sometimes referred to in the art as anti-ligands. As the term receptors is used herein, no difference in meaning is intended. A "Ligand Receptor Pair" is formed when two macromolecules have combined through molecular recognition to form a complex.

Other examples of receptors which can be investigated by this invention include but are not restricted to:

a) Microorganism Receptors

Determination of ligands which bind to receptors, such as specific transport proteins or enzymes essential to survival of microorganisms, is useful for a new class of antibiotics. Of particular value would be antibiotics against opportunistic fungi, protozoa, and those bacteria resistant to the antibiotics in current use.

b) Enzymes

For instance, determining the binding site of enzymes such as the enzymes responsible for cleaving neurotransmitters provides useful information. Determination of ligands which bind to certain receptors to modulate the action of the enzymes which cleave the different neurotransmitters is useful in the development of drugs which can be used in the treatment of disorders of neurotransmission.

c) Antibodies

For instance, the invention may be useful in investigating the ligand-binding site on the antibody molecule which combines with the epitope of an antigen of interest; determining a sequence that mimics an antigenic epitope may lead to the development of vaccines of which the immunogen is based on one or more of such sequences or lead to the development of related diagnostic agents or compounds useful in therapeutic treatments such as for auto-immune diseases (e.g., by blocking the binding of the "self" antibodies). "Antibody" as used herein can also include antibody fragments, such as Fab fragments, which are composed of a light chain and the variable region of a heavy chain.

d) Nucleic Acids

Sequences of nucleic acids may be synthesized to establish DNA or RNA binding sequences.

e) Catalytic Polypeptides

Polymers, preferably polypeptides, which are capable of promoting a chemical reaction involving the conversion of one or more reactants to one or more products. Such polypeptides generally include a binding site specific for at least one reactant or reaction intermediate and an active

US 6,420,169 B1

9

functionality proximate to the binding site, in which the functionality is capable of chemically modifying the bound reactant. Catalytic polypeptides are described in, for example, U.S. application Ser. No. 07/404,920 (now U.S. Pat. No. 5,215,889), which is incorporated herein by reference for all purposes.

f) Hormone Receptors

For instance, the receptors for insulin and growth hormone. Determination of the ligands which bind with high affinity to a receptor is useful in the development of, for example, an oral replacement of the daily injections which diabetics must take to relieve the symptoms of diabetes, and in the other case, a replacement for the scarce human growth hormone which can only be obtained from cadavers or by recombinant DNA technology. Other examples are the vasoconstrictive hormone receptors; determination of those ligands which bind to a receptor may lead to the development of drugs to control blood pressure.

g) Opiate Receptors

Determination of ligands which bind to the opiate receptors in the brain is useful in the development of less-addictive replacements for morphine and related drugs.

8. Substrate

A material having a rigid or semi-rigid surface. In many embodiments, at least one surface of the substrate will be substantially flat, although in some embodiments it may be desirable to physically separate synthesis regions for different polymers with, for example, wells, raised regions, etched trenches, or the like. According to other embodiments, small beads may be provided on the surface which may be released upon completion of the synthesis.

9. Protecting Group

A material which is chemically bound to a monomer unit and which may be removed upon selective exposure to an activator such as electromagnetic radiation and, especially ultraviolet and visible light. Examples of protecting groups with utility herein include those comprising nitropiperonyl, pyrenylmethoxy-carbonyl, nitroveratryl, nitrobenzyl, dimethyl dimethoxybenzyl, 5-bromo-7-nitroindolyl, o-hydroxy- α -methyl cinnamoyl, and 2-oxymethylene anthraquinone.

10. Predefined Region

A predefined region is a localized area on a surface which is, was, or is intended to be activated for formation of a polymer. The predefined region may have any convenient shape, e.g., circular, rectangular, elliptical, wedge-shaped, etc. For the sake of brevity herein, "predefined regions" are sometimes referred to simply as "regions."

11. Substantially Pure

A polymer is considered to be "substantially pure" within a predefined region of a substrate when it exhibits characteristics that distinguish it from other predefined regions. Typically, purity will be measured in terms of biological activity or function as a result of uniform sequence. Such characteristics will typically be measured by way of binding with a selected ligand or receptor. Preferably the region is sufficiently pure such that the predominant species in the predefined region is the desired sequence. According to preferred aspects of the invention, the polymer is 5% pure, more preferably more than 10% pure, preferably more than 20% pure, more preferably more than 80% pure, more preferably more than 90% pure, more preferably more than 95% pure, where purity for this purpose refers to the ratio of the number of ligand molecules formed in a predefined region having a desired sequence to the total number of molecules formed in the predefined region.

12. Activator

10

An energy source adapted to render a group active and which is directed from a source to a predefined location on a substrate. A primary illustration of an activator is light. Other examples of activators include ion beams, electric fields, magnetic fields, electron beams, x-ray, and the like.

13. Combinatorial Synthesis Strategy

An ordered strategy for parallel synthesis of diverse polymer sequences by sequential addition of reagents which may be represented by a reactant matrix, and a switch matrix, the product of which is a product matrix. A reactant matrix is a 1 column by m row matrix of the building blocks to be added. The switch matrix is all or a subset of the binary numbers, preferably ordered, between 1 and m arranged in columns. In preferred embodiments, a "binary strategy" is one in which at least two successive steps illuminate half of a region of interest on the substrate. In a binary synthesis strategy, all possible compounds which can be formed from an ordered set of reactants are formed. In most preferred embodiments, binary synthesis refers to a synthesis strategy which also factors a previous addition step. For example, a strategy in which a switch matrix for a masking strategy halves regions that were previously illuminated, illuminating about half of the previously illuminated region and protecting the remaining half (while also protecting about half of previously protected regions and illuminating about half of previously protected regions). It will be recognized that binary rounds may be interspersed with non-binary rounds and that only a portion of a substrate may be subjected to a binary scheme. A combinatorial "masking" strategy is a synthesis which uses light or other deprotecting or activating agents to remove protecting groups from materials for addition of other materials such as amino acids. In some embodiments, selected columns of the switch matrix are arranged in order of increasing binary numbers in the columns of the switch matrix.

14. Linker

A molecule or group of molecules attached to a substrate and spacing a synthesized polymer from the substrate for exposure/binding to a receptor.

15. Abbreviations

The following abbreviations are intended to have the following meanings:

BOC:	benzyloxycarbonyl.
BOP:	benzotriazol-1-yloxytris-(dimethylamino). phosphonium hexafluorophosphate.
CCD:	charge coupled device.
DCC:	dicyclohexylcarbodiimide.
DCM:	dichloromethane; methylene chloride.
DDZ:	dimethoxydimethylbenzyl.
DIEA:	N,N-diisopropylethylamine.
DMAP:	4-dimethylaminopyridine.
DMF:	dimethyl formamide.
DMT:	dimethoxytrityl.
FMOC:	fluorenylmethyloxycarbonyl.
HBTU:	2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate.
HOBt:	1-hydroxybenzotriazole hydrate.
NMP:	N-methylpyrrolidone.
NV:	nitroveratryl.
NVOC:	6-nitroveratryloxycarbonyl.
PG:	protective group.
TFA:	trifluoroacetic acid.
THF:	tetrahydrofuran.

II. General

The present invention provides synthetic strategies and devices for the creation of large scale chemical diversity. Solid-phase chemistry, photolabile protecting groups, and

US 6,420,169 B1

11

photolithography are brought together to achieve light-directed spatially-addressable parallel chemical synthesis in preferred embodiments.

The invention is described herein for purposes of illustration primarily with regard to the preparation of peptides and nucleotides, but could readily be applied in the preparation of other polymers. Such polymers include, for example, both linear and cyclic polymers of nucleic acids, polysaccharides, phospholipids, and peptides having either .alpha.-, .beta.-, or .omega.-amino acids, heteropolymers in which a known drug is covalently bound to any of the above, polyurethanes, polyesters, polycarbonates, polyureas, polyamides, polyethyleneimines, polyarylene sulfides, polysiloxanes, polyimides, polyacetates, or other polymers which will be apparent upon review of this disclosure. It will be recognized further, that illustrations herein are primarily with reference to C- to N-terminal synthesis, but the invention could readily be applied to N- to C-terminal synthesis without departing from the scope of the invention. Methods for forming cyclic and reversed polarity peptides and other polymers are described in U.S. Pat. No. 5,242,974 and previously incorporated herein by reference.

The prepared substrate may, for example, be used in screening a variety of polymers as ligands for binding with a receptor, although it will be apparent that the invention could be used for the synthesis of a receptor for binding with a ligand. The substrate disclosed herein will have a wide variety of other uses. Merely by way of example, the invention herein can be used in determining peptide and nucleic acid sequences which bind to proteins, finding sequence-specific binding drugs, identifying epitopes recognized by antibodies, and evaluation of a variety of drugs for clinical and diagnostic applications, as well as combinations of the above.

The invention preferably provides for the use of a substrate "S" with a surface. Linker molecules "L" are optionally provided on a surface of the substrate. The purpose of the linker molecules, in some embodiments, is to facilitate receptor recognition of the synthesized polymers.

Optionally, the linker molecules may be chemically protected for storage purposes. A chemical storage protecting group such as t-BOC (t-butoxycarbonyl) may be used in some embodiments. Such chemical protective groups would be chemically removed upon exposure to, for example, acidic solution and would serve to protect the surface during storage and be removed prior to polymer preparation.

On the substrate or a distal end of the linker molecules, a functional group with a protecting group P_0 is provided. The protecting group P_0 may be removed upon exposure to radiation, electric fields, electric currents, or other activators to expose the functional group.

In a preferred embodiment, the radiation is ultraviolet (UV), infrared (IR), or visible light. As more fully described below, the protecting group may alternatively be an electrochemically-sensitive group which may be removed in the presence of an electric field. In still further alternative embodiments, ion beams, electron beams, or the like may be used for deprotection.

A. Deprotection and Addition

Concurrently or after exposure of a known region of the substrate to light, the surface is contacted with a first monomer unit M_1 which reacts with the functional group which has been exposed by the deprotection step. The first monomer includes a protecting group P_1 . P_1 may or may not be the same as P_0 .

Accordingly, after a first cycle, known first regions of the surface may comprise the sequence:

12

$$S-L-M_1-P_1$$

while remaining regions of the surface comprise the sequence:

$$S-L-P_0.$$

Thereafter, second regions of the surface (which may include the first region) are exposed to light and contacted with a second monomer M_2 (which may or may not be the same as M_1) having a protecting group P_2 . P_2 may or may not be the same as P_0 and P_1 . After this second cycle, different regions of the substrate may comprise one or more of the following sequences:

$$S-L-M_1-M_2-P_2$$

$$S-L-M_2-P_2$$

$$S-L-M_1-P_1$$

and/or

$$S-L-P_0.$$

The above process is repeated until the substrate includes desired polymers of desired lengths. By controlling the locations of the substrate exposed to light and the reagents exposed to the substrate following exposure, the location of each sequence will be known.

Thereafter, the protective groups are removed from some or all of the substrate and the sequences are, optionally, capped with a capping unit C. The process results in a substrate having a surface with a plurality of polymers of the following general formula:

$$S-[L]-(M_i)-(M_j)-(M_k) \dots (M_x)-[C]$$

where square brackets indicate optional groups, and $M_i \dots M_x$ indicates any sequence of monomers. The number of monomers could cover a wide variety of values, but in a preferred embodiment they will range from 2 to 100.

In some embodiments a plurality of locations on the substrate polymers are to contain a common monomer subsequence. For example, it may be desired to synthesize a sequence $S-M_1-M_2-M_3$ at first locations and a sequence $S-M_4-M_2-M_3$ at second locations. The process would commence with irradiation of the first locations followed by contacting with M_1-P , resulting in the sequence $S-M_1-P$ at the first location. The second locations would then be irradiated and contacted with M_4-P , resulting in the sequence $S-M_4-P$ at the second locations. Thereafter both the first and second locations would be irradiated and contacted with the dimer M_2-M_3 , resulting in the sequence $S-M_1-M_2-M_3$ at the first locations and $S-M_4-M_2-M_3$ at the second locations. Of course, common subsequences of any length could be utilized including those in a range of 2 or more monomers, 2 to 100 monomers, 2 to 20 monomers, and a most preferred range of 2 to 3 monomers.

According to other embodiments, a set of masks is used for the first monomer layer and, thereafter, varied light wavelengths are used for selective deprotection. For example, in the process discussed above, first regions are first exposed through a mask and reacted with a first monomer having a first protecting group P_1 , which is removable upon exposure to a first wavelength of light (e.g., IR). Second regions are masked and reacted with a second monomer having a second protective group P_2 which is removable upon exposure to a second wavelength of light (e.g., UV). Thereafter, masks become unnecessary in the

US 6,420,169 B1

13

synthesis because the entire substrate may be exposed alternatively to the first and second wavelengths of light in the deprotection cycle.

The polymers prepared on a substrate according to the above methods will have a variety of uses including, for example, screening for biological activity. In such screening activities, the substrate containing the sequences is exposed to an unlabeled or labeled receptor such as an antibody, receptor on a cell, phospholipid vesicle, or any one of a variety of other receptors. In one preferred embodiment the polymers are exposed to a first, unlabeled receptor of interest and, thereafter, exposed to a labeled receptor-specific recognition element, which is, for example, an antibody. This process will provide signal amplification in the detection stage.

The receptor molecules may bind with one or more polymers on the substrate. The presence of the labeled receptor and, therefore, the presence of a sequence which binds with the receptor is detected in a preferred embodiment through the use of autoradiography, detection of fluorescence with a charge-coupled device, fluorescence microscopy, or the like. The sequence of the polymer at the locations where the receptor binding is detected may be used to determine all or part of a sequence which is complementary to the receptor.

Use of the invention herein is illustrated primarily with reference to screening for biological activity. The invention will, however, find many other uses. For example, the invention may be used in information storage (e.g., on optical disks), production of molecular electronic devices, production of stationary phases in separation sciences, production of dyes and brightening agents, photography, and in immobilization of cells, proteins, lectins, nucleic acids, polysaccharides and the like in patterns on a surface via molecular recognition of specific polymer sequences. By synthesizing the same compound in adjacent, progressively differing concentrations, a gradient will be established to control chemotaxis or to develop diagnostic dipsticks which, for example, titrate an antibody against an increasing amount of antigen. By synthesizing several catalyst molecules in close proximity, more efficient multistep conversions may be achieved by "coordinate immobilization." Coordinate immobilization also may be used for electron transfer systems, as well as to provide both structural integrity and other desirable properties to materials such as lubrication, wetting, etc.

According to alternative embodiments, molecular biodistribution or pharmacokinetic properties may be examined. For example, to assess resistance to intestinal or serum proteases, polymers may be capped with a fluorescent tag and exposed to biological fluids of interest.

FIG. 1 is a flow chart illustrating the process of forming chemical compounds according to one embodiment of the invention. Synthesis occurs on a solid support 2. A pattern of illumination through a mask 4a using a light source 6 determines which regions of the support are activated for chemical coupling. In one preferred embodiment activation is accomplished by using light to remove photolabile protecting groups from selected areas of the substrate.

After deprotection, monomers indicated by "A" in FIG. 1, each bearing a photolabile protecting group (indicated by "X"), are exposed to the surface of the substrate and react with regions that were addressed by light in the preceding step. The substrate is then illuminated through a second mask 4b, which activates another region for reaction with a second protected monomer "B." The process is then repeated using desired masks and mask orientations in

14

combination with selected monomers. The pattern of masks used in these illuminations and the sequence of reactants define the ultimate products and their locations, resulting in diverse sequences at predefined locations, as shown with the sequences ACEG and BDFH in the lower portion of FIG. 1. Preferred embodiments of the invention take advantage of combinatorial masking strategies to form a large number of compounds in a small number of chemical steps.

A high degree of miniaturization is possible because the density of compounds is determined largely with regard to spatial addressability of the activator, in one case the diffraction of light. Each compound is physically accessible and its position is precisely known. Hence, the array is spatially-addressable and its interactions with other molecules can be assessed.

In a particular embodiment shown in FIG. 1, the substrate contains amino groups that are blocked with a photolabile protecting group. Amino acid sequences are made accessible for coupling to a receptor by removal of the photoprotecting groups.

When a polymer sequence to be synthesized is, for example, a polypeptide, amino groups at the ends of linkers attached to a glass substrate are derivatized with, for example, nitroveratryloxycarbonyl (NVOC), a photoremovable protecting group. The linker molecules may be, for example, aryl acetylene, ethylene glycol oligomers containing from 2–10 monomers, diamines, diacids, amino acids, or combinations thereof. Photodeprotection is effected by illumination of the substrate through, for example, a mask wherein the pattern has transparent regions with dimensions of, for example, less than 1 cm^2 , 10^{-1} cm^2 , 10^{-2} cm^2 , 10^{-3} cm^2 , 10^{-4} cm^2 , 10^{-5} cm^2 , 10^{-6} cm^2 , 10^{-7} cm^2 , 10^{-8} cm^2 , or 10^{-10} cm^2 . In a preferred embodiment, the regions are between about $10\times 10\text{ }\mu\text{m}$ and $500\times 500\text{ }\mu\text{m}$. According to some embodiments, the masks are arranged to produce a checkerboard array of polymers, although any one of a variety of geometric configurations may be utilized.

1. EXAMPLE

Polymer Synthesis

FIG. 2 illustrates one embodiment of the invention disclosed herein in which a substrate 8 is shown in cross-section. Essentially, any conceivable substrate may be employed in the invention. The substrate may be biological, nonbiological, organic, inorganic, or a combination of any of these, existing as particles, strands, precipitates, gels, sheets, tubing, spheres, containers, capillaries, pads, slices, films, plates, slides, etc. The substrate may have any convenient shape, such as a disc, square, sphere, circle, etc. The substrate is preferably flat but may take on a variety of alternative surface configurations. For example, the substrate may contain raised or depressed regions on which the synthesis takes place. The substrate and its surface preferably form a rigid support on which to carry out the reactions described herein. The substrate and its surface are also chosen to provide appropriate light-absorbing characteristics. For instance, the substrate may be a polymerized Langmuir Blodgett film, functionalized glass, Si, Ge, GaAs, GaP, SiO_2 , SiN_4 , modified silicon, any of the different crystal lattices made with silicon or gallium arsenide that are commercially available and used in semiconductor manufacturing, or any one of a wide variety of gels or polymers such as (poly)tetrafluoroethylene, (poly)vinylidenedifluoride, polystyrene, polycarbonate, or combinations thereof. Other substrate materials will be readily apparent to those of skill in the art upon review of this disclosure. In a preferred embodiment the substrate is flat glass or single-crystal silicon with surface relief features of less than $10\text{ }\text{\AA}$.

US 6,420,169 B1

15

According to some embodiments, the surface of the substrate is etched using well known techniques to provide for desired surface features. For example, by way of the formation of trenches, v-grooves, mesa structures, or the like, the synthesis regions may be more closely placed within the focus point of impinging light, be provided with reflective "mirror" structures for maximization of light collection from fluorescent sources, or the like.

Surfaces on the solid substrate will usually, though not always, be composed of the same material as the substrate. Thus, the surface may be composed of any of a wide variety of materials, for example, polymers, plastics, resins, polysaccharides, silica or silica-based materials, carbon, metals, inorganic glasses, membranes, or any of the above-listed substrate materials. In some embodiments the surface may provide for the use of caged binding members which are attached firmly to the surface of the substrate. Preferably, the surface will contain reactive groups, which could be carboxyl, amino, hydroxyl, or the like. Most preferably, the surface will be optically transparent and will have surface Si—OH functionalities, such as are found on silica surfaces.

The surface **10** of the substrate is preferably provided with a layer of linker molecules **12**, although it will be understood that the linker molecules are not required elements of the invention. The linker molecules are preferably of sufficient length to permit polymers in a completed substrate to interact freely with molecules exposed to the substrate. The linker molecules should be 6–50 atoms long to provide sufficient exposure. The linker molecules may be, for example, aryl acetylene, ethylene glycol oligomers containing 2–10 monomer units, diamines, diacids, amino acids, or combinations thereof. Other linker molecules may be used in light of this disclosure.

According to alternative embodiments, the linker molecules are selected based upon their hydrophilic/hydrophobic properties to improve presentation of synthesized polymers to certain receptors. For example, in the case of a hydrophilic receptor, hydrophilic linker molecules will be preferred so as to permit the receptor to more closely approach the synthesized polymer.

According to another alternative embodiment, linker molecules are also provided with a photocleavable group at an intermediate position. The photocleavable group is preferably cleavable at a wavelength different from the protective group. This enables removal of the various polymers following completion of the synthesis by way of exposure to the different wavelengths of light. One can also modify the linker molecule with a photocleavable group, which, when removed, will induce a conformational change in the polymer attached to the linker.

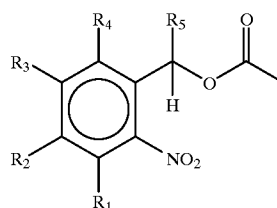
The linker molecules can be attached to the substrate via carbon-carbon bonds using, for example, (poly) trifluorochloroethylene surfaces, or preferably, by siloxane bonds (using, for example, glass or silicon oxide surfaces). Siloxane bonds with the surface of the substrate may be formed in one embodiment via reactions of linker molecules bearing trichlorosilyl groups. The linker molecules may optionally be attached in an ordered array, i.e., as parts of the head groups in a polymerized Langmuir Blodgett film. In alternative embodiments, the linker molecules are adsorbed to the surface of the substrate.

The linker molecules and monomers used herein are provided with a functional group to which is bound a protective group. Preferably, the protecting group is on the distal or terminal end of the linker molecule opposite the substrate. The protecting group may be either a negative protecting group (i.e., the protecting group renders the linker

16

molecules less reactive with a monomer upon exposure) or a positive protecting group (i.e., the protecting group renders the linker molecules more reactive with a monomer upon exposure). In the case of negative protective groups an additional step of reactivation will be required. In some embodiments, this will be done by heating. Those of skill in the art will also note that more than one functional group can be employed on either the linker or the monomer, i.e., to facilitate the synthesis of branched or "dendritic" structures.

The protecting group on the linker molecules may be selected from a wide variety of positive light-reactive groups preferably including nitro aromatic compounds such as o-nitrobenzyl derivatives or benzylsulfonyl. In a preferred embodiment, 6-nitroveratryloxycarbonyl (NVOC), 2-nitrobenzyloxycarbonyl (NBOC) or α,α -dimethyldimethoxybenzyloxycarbonyl (DDZ) is used. In one embodiment, a nitro aromatic compound containing a benzylic hydrogen ortho to the nitro group is used, i.e., a chemical of the form:



where R_1 is alkoxy, alkyl, halo, aryl, alkenyl, or hydrogen; R_2 is alkoxy, alkyl, halo, aryl, nitro, or hydrogen; R_3 is alkoxy, alkyl, halo, nitro, aryl, or hydrogen; R_4 is alkoxy, alkyl, hydrogen, aryl, halo, or nitro; and R_5 is alkyl, alkynyl, cyano, alkoxy, hydrogen, halo, aryl, or alkenyl. Other materials which may be used include o-hydroxy- α -methyl cinamoyl derivatives.

In an alternative embodiment the positive reactive group is activated for reaction with reagents in solution. For example, a 5-bromo-7-nitro indoline group, when bound to a carbonyl, undergoes reaction upon exposure to light at 420 nm.

In a second alternative embodiment, the reactive group on the linker molecule is selected from a wide variety of negative light-reactive groups including a cinammate group.

Alternatively, the reactive group is activated or deactivated by electron beam lithography, x-ray lithography, or any other radiation. Suitable reactive groups for electron beam lithography include sulfonyl. Other methods may be used including, for example, exposure to a current source. Other reactive groups and methods of activation may be used in light of this disclosure.

As shown in FIG. 2, the linking molecules are preferably exposed to, for example, irradiation, such as light, through a suitable mask **14** using photolithographic techniques of the type known in the semiconductor industry and described in, for example, Sze, *VLSI Technology*, McGraw-Hill (1983), and Mead et al., *Introduction to VLSI Systems*, Addison-Wesley (1980), which are incorporated herein by reference for all purposes. The light may be directed at either the surface containing the protective groups or at the back of the substrate, so long as the substrate is transparent to the wavelength of light needed for removal of the protective groups. In the embodiment shown in FIG. 2, light is directed at the surface of the substrate containing the protective groups. FIG. 2 illustrates the use of such masking techniques as they are applied to a positive reactive group so as to activate linking molecules and expose functional groups in areas **16a** and **16b**.

The mask 14 is in one embodiment a transparent support material selectively coated with a layer of opaque material. Portions of the opaque material are removed, leaving opaque material in the precise pattern desired on the substrate surface. The mask is brought into close proximity with, imaged on, or brought directly into contact with the substrate surface as shown in FIG. 2. "Openings" in the mask correspond to locations on the substrate where it is desired to remove photoremovable protective groups from the substrate. Alignment may be performed using conventional alignment techniques in which alignment marks (not shown) are used to accurately overlay successive masks with previous patterning steps, or more sophisticated techniques may be used. For example, interferometric techniques such as the one described in Flanders et al., "A New Interferometric Alignment Technique," *App. Phys. Lett.* (1977) 31:426-428, which is incorporated herein by reference, may be used.

To enhance contrast of light applied to the substrate, it is desirable to provide contrast enhancement materials between the mask and the substrate according to some embodiments. This contrast enhancement layer may comprise a molecule which is decomposed by light such as quinone diazide or a material which is transiently bleached at the wavelength of interest. Transient bleaching of materials will allow greater penetration where light is applied, thereby enhancing contrast. Alternatively, contrast enhancement may be provided by way of a clad fiber optic bundle.

The light may be from a conventional incandescent source, a laser, a laser diode, or the like. If non-collimated sources of light are used it may be desirable to provide a thick- or multi-layered mask to prevent spreading of the light onto the substrate. It may, further, be desirable in some embodiments to utilize groups which are sensitive to different wavelengths to control synthesis. For example, by using groups which are sensitive to different wavelengths, it is possible to select branch positions in the synthesis of a polymer or eliminate certain masking steps. Several reactive groups along with their corresponding wavelengths for deprotection are provided in Table 1.

TABLE 1

Group	Approximate Deprotection Wavelength
Nitroveratryloxy carbonyl (NVOC)	UV (300-400 nm)
Nitrobenzyloxy carbonyl (NBOC)	UV (300-350 nm)
Dimethyl dimethoxybenzyloxy carbonyl	UV (280-300 nm)
5-Bromo-7-nitroindolyl	UV (420 nm)
o-Hydroxy- α -methyl cinnamoyl	UV (300-350 nm)
2-Oxymethylene anthraquinone	UV (350 nm)

Note that different photoprotected monomers, such as amino acids, can exhibit different photolysis rates. See, for example, "The Peptides, Analysis, Synthesis, Biology" Chapter 8, E. Gross and J. Meienhofer, Eds., Academic Press, Inc. (1980); and PCT application WO 89/10931. It may be desirable to utilize photoprotected monomers with substantially similar photolysis rates in a particular application. To obtain such a set of photoprotected monomers, one merely needs to select the appropriate photoprotecting group for each monomer in the set. In similar fashion, one can prepare a set of photoprotected monomers with substantially different photolysis rates (from monomer to monomer) by appropriate choice of photoprotecting groups.

While the invention is illustrated primarily herein by way of the use of a mask to illuminate selected regions the substrate, other techniques may also be used. For example,

the substrate may be translated under a modulated laser or diode light source. Such techniques are discussed in, for example, U.S. Pat. No. 4,719,615 (Feyrer et al.), which is incorporated herein by reference. In alternative embodiments a laser galvanometric scanner is utilized. In other embodiments, the synthesis may take place on or in contact with a conventional liquid crystal (referred to herein as a "light valve") or fiber optic light sources. By appropriately modulating liquid crystals, light may be selectively controlled so as to permit light to contact selected regions of the substrate. Alternatively, synthesis may take place on the end of a series of optical fibers to which light is selectively applied. Other means of controlling the location of light exposure will be apparent to those of skill in the art.

The substrate may be irradiated either in contact or not in contact with a solution and is, preferably, irradiated in contact with a solution. The solution contains reagents to prevent the by-products formed by irradiation from interfering with synthesis of the polymer according to some embodiments. Such by-products might include, for example, carbon dioxide, nitrosocarbonyl compounds, styrene derivatives, indole derivatives, and products of their photochemical reactions. Alternatively, the solution may contain reagents used to match the index of refraction of the substrate. Reagents added to the solution may further include, for example, acidic or basic buffers, thiols, substituted hydrazines and hydroxylamines, reducing agents (e.g., NADH) or reagents known to react with a given functional group (e.g., aryl nitroso+glyoxylic acid→aryl formhydroxamate+CO₂).

Either concurrently with or after the irradiation step, the linker molecules are washed or otherwise contacted with a first monomer, illustrated by "A" in regions 18a and 18b in FIG. 3. The first monomer reacts with the activated functional groups of the linker molecules which have been exposed to light. The first monomer, which is preferably an amino acid, is also provided with a photoprotective group. The photoprotecting group on the monomer may be the same as or different than the protecting group used in the linker molecules, and may be selected from any of the above-described protective groups. In one embodiment, the protective groups for the A monomer is selected from the group NBOC and NVOC.

As shown in FIG. 4, the process of irradiating is thereafter repeated, with a mask repositioned so as to remove linkage protective groups and expose functional groups in regions 20a and 20b which are illustrated as being regions which were protected in the previous masking step. As an alternative to repositioning of the first mask, in many embodiments a second mask will be utilized. In other alternative embodiments, some steps may provide for illuminating a common region in successive steps. As shown in FIG. 4, it may be desirable to provide separation between irradiated regions. For example, separation of about 1-5 μ m may be appropriate to account for alignment tolerances.

As shown in FIG. 5, the substrate is then exposed to a second protected monomer "B," producing B regions 22a and 22b. Thereafter, the substrate is again masked so as to remove the protective groups and expose reactive groups on A region 18a and B region 18b. The substrate is again exposed to monomer B, resulting in the formation of the structure shown in FIG. 6. The dimers B-A and B-B have been produced on the substrate.

A subsequent series of masking and contacting steps similar to those described above with A (not shown) provides the structure shown in FIG. 7. The process provides all possible dimers of B and A, i.e., B-A, A-B, A-A, and B-B.

The substrate, the area of synthesis, and the area for synthesis of each individual polymer could be of any size or shape. For example, squares, ellipsoids, rectangles, triangles, circles, or portions thereof, along with irregular geometric shapes, may be utilized. Duplicate synthesis areas may also be applied to a single substrate for purposes of redundancy.

In some embodiments a single substrate supports more than about 10 different monomer sequences and preferably more than about 100 different monomer sequences, although in some embodiments more than about 10³, 10⁴, 10⁵, 10⁶, 10⁷, or 10⁸ different sequences are provided on a substrate. of course, within a region of the substrate in which a monomer sequence is synthesized, it is preferred that the monomer sequence be substantially pure. In some embodiments, regions of the substrate contain polymer sequences which are at least about 1%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, or 99% pure.

According to some embodiments, several sequences are intentionally provided within a single region so as to provide an initial screening for biological activity, after which materials within regions exhibiting significant binding are further evaluated.

2. EXAMPLE

It is important to achieve good contrast between exposed and non-exposed regions on a substrate. Otherwise, unwanted products will be formed as, for example, when a monomer is added to a polymer in an unexposed or dark region. It has been determined that long exposure times will often result in more complete monomer coupling in activated regions. However, if the substrate is exposed for too long a period, the photolysis near the exposure edges will decrease, i.e., the contrast between exposed and unexposed regions will be reduced.

The deprotection of an NVOC-protected amine group was employed to model masking resolution.

The deprotection of the NVOC protected amine by ultra-violet light is a first order reaction of reactant A being converted to product B:



dA/dt = kA

ln A/A0 = ln A0/A0 = -kT

A/A0 = e^{-kt} = % NVOC protected amines after time t

1 - A/A0 = 1 - e^{-kt} = % free amines at t

where k is rate of photolysis=(physical constants)×I, and I is light intensity at 365 nm=13 mw/cm². A is the concentration of the reactant and A₀ is that concentration at t=0.

The dark areas were modelled as if they were being photolyzed with a fraction of the light intensity and a new rate of photolysis to the areas beyond the photolysis site was defined.

I'=CI

k'=Ck

r = (1 - A'/A0) = FA/FB = 1 - e^{-k't}
k' = -ln(1 - r)/t

F is the height of the histogram used to analyze the experimental results. Thus, F is also the florescence intensity at a distance from the edge of photolysis.

Contrast was investigated by photolysis through a binary mask (12800 μm×6400 μm) for 660, 1320, and 9990 seconds. The contrast ratio was measured (as function of distance from the photolysis edge) as the ratio of the work height of the histogram in the dark area to the height of the histogram in the light area, i.e. FA/FB. The results are presented below.

N	time	0 μM	50 μm	100 μm	200 μm
1	660 sec.	1	0.4	0.1	0.01
2	1320 sec.	1	0.7	0.3	0.03
3	9900 sec.	1	0.92	0.7	0.39

Thus, it can be seen that the photolysis fidelity is a function of both the time of exposure and the distance from the edge.

3. EXAMPLE

Slide Preparation

Before attachment of reactive groups it is preferred to clean the substrate which is, in a preferred embodiment, a glass substrate such as a microscope slide or cover slip. According to one embodiment the slide is soaked in an alkaline bath consisting of, for example, 1 liter of 95% ethanol with 120 ml of water and 120 grams of sodium hydroxide for 12 hours. The slides are then washed under running water and allowed to air dry, and rinsed once with a solution of 95% ethanol.

The slides are then aminated with, for example, aminopropyltriethoxysilane for the purpose of attaching amino groups to the glass surface on linker molecules, although any omega functionalized silane could also be used for this purpose. In one embodiment 0.1% aminopropyltriethoxysilane is utilized, although solutions with concentrations from 10⁻⁷% to 10% may be used, with about 10⁻³% to 2% preferred. A 0.1% mixture is prepared by adding to 100 ml of a 95% ethanol/5% water mixture, 100 microliters (μl) of aminopropyltriethoxysilane. The mixture is agitated at about ambient temperature on a rotary shaker for about 5 minutes. 500 μl of this mixture is then applied to the surface of one side of each cleaned slide. After 4 minutes, the slides are decanted of this solution and rinsed three times by dipping in, for example, 100% ethanol.

After the plates dry, they are placed in a 110–120° C. vacuum oven for about 20 minutes, and then allowed to cure at room temperature for about 12 hours in an argon environment. The slides are then dipped into DMF (dimethylformamide) solution, followed by a thorough washing with dichloromethane.

The aminated surface of the slide is then exposed to about 500 μl of, for example, a 30 millimolar (mM) solution of NVOC-GABA (gamma amino butyric acid) NHS (N-hydroxysuccinimide) in DMF for attachment of a NVOC-GABA to each of the amino groups.

The surface is washed with, for example, DMF, dichloromethane, and ethanol.

US 6,420,169 B1

21

Any unreacted aminopropyl silane on the surface—that is, those amino groups which have not had the NVOC-GABA attached—are now capped with acetyl groups (to prevent further reaction) by exposure to a 1:3 mixture of acetic anhydride in pyridine for 1 hour. Other materials which may perform this residual capping function include trifluoroacetic anhydride, formicacetic anhydride, or other reactive acylating agents. Finally, the slides are washed again with DMF, methylene chloride, and ethanol.

4. EXAMPLE

Synthesis of a Dimer of an Aminopropyl Group and a Fluorescent Group

In synthesizing the dimer of an aminopropyl group and a fluorescent group, a functionalized Durapore™ membrane was used as a substrate. The Durapore™ membrane was a polyvinylidene difluoride with aminopropyl groups. The aminopropyl groups were protected with the DDZ group by reaction of the carbonyl chloride with the amino groups, a reaction readily known to those of skill in the art. The surface bearing these groups was placed in a solution of THF and contacted with a mask bearing a checkerboard pattern of 1 mm opaque and transparent regions. The mask was exposed to ultraviolet light having a wavelength down to at least about 280 nm for about 5 minutes at ambient temperature, although a wide range of exposure times and temperatures may be appropriate in various embodiments of the invention. For example, in one embodiment, an exposure time of between about 1 and 5000 seconds may be used at process temperatures of between -70 and +50° C.

In one preferred embodiment, exposure times of between about 1 and 500 seconds at about ambient pressure are used. In some preferred embodiments, pressure above ambient is used to prevent evaporation.

The surface of the membrane was then washed for about 1 hour with a fluorescent label which included an active ester bound to a chelate of a lanthanide. Wash times will vary over a wide range of values from about a few minutes to a few hours. These materials fluoresce in the red and the green visible region. After the reaction with the active ester in the fluorophore was complete, the locations in which the fluorophore was bound could be visualized by exposing them to ultraviolet light and observing the red and the green fluorescence. It was observed that the derivatized regions of the substrate closely corresponded to the original pattern of the mask.

5. EXAMPLE

Removal of NVOC and Attachment of a Fluorescent Marker

NVOC-GABA groups were attached as described above, except that the substrate was a glass slide. The entire surface of one slide was exposed to light so as to expose a free amino group at the end of the gamma amino butyric acid. This slide, and a duplicate which was not exposed, were then exposed to fluorescein isothiocyanate (FITC).

FIG. 8A illustrates the slide which was not exposed to light, but which was exposed to FITC. Fluorescence on the surface was measured by excitation using 488 nm laser light and photomultiplier detection through appropriate fluorescein emission filters described in additional detail below. The units of the x axis are time (msec) and the units of the y axis are counts. The trace contains a certain amount of background fluorescence. The duplicate slide was exposed to 350 nm broadband illumination for about 1 minute (12 mW/cm², -350 nm illumination), washed and reacted with FITC. The fluorescence curves for this slide are shown in FIG. 8B. A large increase in the level of fluorescence is observed, which indicates photolysis has exposed a number

22

of amino groups on the surface of the slides for attachment of a fluorescent marker.

6. EXAMPLE

Use of a Mask in Removal of NVOC

The next experiment was performed with a 0.1% amino-propylated slide. Light from a Hg—Xe arc lamp was imaged onto the substrate through a laser-ablated chrome-on-glass mask in direct contact with the substrate.

This slide was illuminated for approximately 5 minutes, with 12 mW of 350 nm broadband light and then reacted with the 1 mM FITC solution. It was put on the laser detection scanning stage and a graph was plotted as a two-dimensional representation of position versus fluorescence intensity. The fluorescence intensity (in counts) as a function of location is given on the scale to the right of FIG. 9A for a mask having 100×100 μm squares.

The experiment was repeated a number of times through various masks. The fluorescence pattern for a 50 μm mask is illustrated in FIG. 9B, for a 20 μm mask in FIG. 9C, and for a 10 μm mask in FIG. 9D. The mask pattern is distinct down to at least about 10 μm squares using this lithographic technique.

7. EXAMPLE

In one example of the invention, free amino groups were fluorescently labelled by treatment of the entire substrate surface with fluorescein isothiocyanate (FITC) after photodeprotection. Glass microscope slides were cleaned, aminated by treatment with 0.1% aminopropyltriethoxysilane in 95% ethanol, and incubated at 110° C. for 20 min. The aminated surface of the slide was then exposed to a 30 mM solution of the N-hydroxysuccinimide ester of NVOC-GABA (nitroveratryloxycarbonyl- τ -amino butyric acid) in DMF. The NVOC protecting group was photolytically removed by imaging the 365 nm output from a Hg arc lamp through a chrome on glass 100 μm checkerboard mask onto the substrate for 20 min at a power density of 12 mW/cm². The exposed surface was then treated with 1 mM FITC in DMF. The substrate surface was scanned in an epifluorescence microscope (Zeiss Axioskop 20) using 488 nm excitation from an argon ion laser (Spectra-Physics model 2025). The fluorescence emission above 520 nm was detected by a cooled photomultiplier (Hamamatsu 943-02) operated in a photon counting mode. Fluorescence intensity was translated into a color display with red in the highest intensity and black in the lowest intensity areas. The presence of a high-contrast fluorescent checkerboard pattern of 100×100 μm elements revealed that free amino groups were generated in specific regions by spatially-localized photodeprotection.

8. EXAMPLE

Slide preparation is illustrated below. Slides used in synthesis may be detergent cleaned, glass slides. Such glass slides may be, for example, 1"×3" smooth cut, 0.7 mm thick, anti-scratch coated, or 2"×3" smooth cut, 0.7 mm thick, anti-scratch coated from Erie Scientific. The slides are soaked in 10% Micro™ detergent (from Baxter), individually scrubbed, and immersed in deionized H₂O until all slides have been scrubbed. The slides are then subjected to 10 minute sonication in 70° C. "Micro" detergent and rinsed 10× with deionized H₂O. This process is followed by a 3 minute immersion in 70° C. 10% (w/v) NaOH. The slides are then rinsed 10× with deionized H₂O, followed by a 1 minute immersion in 1% HCl. The slides are then again rinsed 10× with deionized H₂O, followed by another 10 minute sonication in 70° C. deionized water and are rinsed

US 6,420,169 B1

23

3–4× in deionized H₂O. The slides are then ethanol rinsed and dried with nitrogen or argon. The slides are then inspected visually for spots and scratches, preferably in a yellow light with a black background.

Alternatively or in addition, the slides are acid cleaned. The slides are loaded into teflon racks and subjected to a 30 minute immersion in Nochromix™ (Aldrich) solution with 36 g per liter of concentrated H₂SO₄, which is regenerated if discolored, filtered (glass fiber filter) to remove particulate matter, and provided with occasional agitation. The slides are then rinsed for 1 min. in deionized H₂O with vigorous agitation. The slides are then placed for 10 minutes in a rinse tank with 14 psi argon or nitrogen bubbling, a full open deionized water tap, and occasional agitation.

The slides are then immersed for 3 minutes in 70° C. 10%(w/v) NaOH, followed by a 1 minute rinse deionized H₂O with vigorous agitation, followed by 10 minutes in a rinse tank. The slides are then immersed for 1 min. in 1% HCl, and rinsed for 5 minutes in a rinse tank. The slides are then ethanol rinsed, dried with nitrogen or argon, and inspected visually for spots and scratches.

tBOC aminopropyl derivatization is illustrated below. The slides are loaded into plastic staining jars. Preferably the slides are completely dry, with 9 slides per jar. Silation reagents are then mixed as follows:

a pre-mix 1:10 mole ratio of tBOCaminopropyltriethoxysilane:methyltriethoxysilane

tBOC-aminopropyltriethoxysilane:

MW=321.49

d=0.945 g/ml; and

methyltriethoxysilane:

MW=178.30

d=0.895 g/ml

1:10 ratio=1 ml tBOC-aminopropyl to 5.86 ml methyltriethoxysilane

The reagents are kept anhydrous and stored under argon. The silation reagent is diluted to 1%(v/v) in dichloromethane (DCM), mixed well, and 60 ml per jar are added. The jars are capped and left overnight.

The silation solution is poured into a plastic container, rinsed with dichloromethane (DCM), and the slides are rinsed with toluene. The toluene is then poured off, and the slides are dried immediately with argon. The slides are loaded into glass drying racks, inspected for streaks, and allowed to stand for approximately 30 minutes.

The slides are then baked for 1 hour in 100° C. oven with the glass racks in metal trays covered with foil. The oven is preferably no hotter than 110° C. The slides are then cooled and numbered using an engraving tool.

Aminocaproic acid coupling is illustrated below. The tBOC-aminopropyl slides are loaded into glass staining jars with 15 slides per jar. The slides are then deprotected and neutralized with a 30 minute immersion in 50% TFA/DCM, a rinse for 2 minutes in DCM, and a rinse 2× in 5% DIEA/DCM for 5 minutes each, followed by a rinse with dichloromethane, and a rinse with ethanol. The slides are then dried with argon, and derivatized within one hour.

The volume of solution necessary is equal to 0.4 ml×# slides. The concentration of the solution is 100 mM NVOC-aminocaproic acid, 110 mM HOBt, 200 mM DIEA, and 100 mM BOC.

The slides are placed face up on glass plates, and 0.4 ml solution is layered per slide. The slides are then covered in plastic trays and allowed to sit for 2 hours. The slides are then rinsed with DMF or NMP, rinsed with DCM, and rinsed with EtOH.

24

The slides are capped with acetic anhydride by immersion for 1 hour in 25% acetic anhydride/pyridine and 0.1% DMAP. The slides are then rinsed with DMF or NMP, rinsed with DCM, and rinsed with EtOH. The slides are then dried with argon and stored in a light-tight container.

Biotinylation of NVOC-aminocaproic slides may be desirable in some instances (for example, see infra Section V.C.) and is achieved as follows. The slides are photolyzed in 5 mM H₂SO₄/dioxane with appropriate masking and a large area mercury illuminator with a 350–450 nm dichroic reflector and a 12 minute exposure at 12–13 mW/cm².

The slides are collected in dioxane until all slides have been exposed, washed 2× in 5% DIEA/DMF for 5 minutes each, and rinsed with DMF, DCM, and EtOH. The slides are then dried with argon, and are preferably derivatized within one hour.

The volume of solution necessary is equal to 0.4 ml×# slides. The concentration of the solutions is 100 mM Biotin, 110 mM HOBt, and 200 mM DIEA and 100 mM BOC. A heat gun is used to help dissolve biotin. 100 mM BOP (MW=442.29) is dissolved in 1/10th final volume NMP or DMF. The solutions are mixed, capped, and allowed to stand 10 minutes. The final volume is adjusted with NMP or DMF and the solutions are mixed well.

The slides are placed face up on glass plates, and 0.4 ml solution per slide is layered onto the slide. The slides are covered in plastic trays, allowed to sit for 2 hours, and rinsed with NMP or DMF, then rinsed with DCM, then rinsed with EtOH, and dried with argon.

FITC labeling of amines is achieved as follows. First the amines are deprotected and neutralized by photolyzing NVOC-aminocaproic slides. The photolysis takes place in 5 mM H₂SO₄/dioxane with appropriate masking using a large area illuminator with a 350–450 nm dichroic reflector, 12–13 mW/cm², and a 12 minute exposure. The slides are rinsed in dioxane, and then rinsed 2× in 5% DIEA/DMF; 5 minutes each. The slides are then rinsed with DMF.

Deprotected amines are labeled by immersing slides in imM FITC/DMF for 1 hour, rinsing with DMF, rinsing with DCM, rinsing with ethanol, and drying slides with argon.

B. Antibody Recognition

In one preferred embodiment the substrate is used to determine which of a plurality of amino acid sequences is recognized by an antibody of interest.

For purposes of increasing the signal-to-noise ratio of the technique, some embodiments of the invention provide for exposure of the substrate to a first labeled or unlabeled receptor followed by exposure of a labeled, second receptor (e.g., an antibody) which binds at multiple sites on the first receptor. If, for example, the first receptor is an antibody derived from a first species of an animal, the second receptor is an antibody derived from a second species directed to epitopes associated with the first species. In the case of a mouse antibody, for example, fluorescently labeled goat antibody or antiserum which is antimouse may be used to bind at multiple sites on the mouse antibody, providing several times the fluorescence compared to the attachment of a single mouse antibody at each binding site. This process may be repeated again with additional antibodies (e.g., goat-mouse-goat, etc.) for further signal amplification.

1. EXAMPLE

Attachment of YGGFL and Subsequent Exposure to Herz Antibody and Goat Antimouse

In order to establish that receptors to a particular polypeptide sequence would bind to a surface-bound peptide and be detected, t-BOC protected Leu enkephalin was coupled to the surface and recognized by an antibody. A slide was

US 6,420,169 B1

25

derivatized with 0.1% amino propyl-triethoxysilane and protected with NVOC-6-amino caproic acid or NVOC-GABA. A 500 μm checkerboard mask was used to expose the slide in a flow cell using backside contact printing. The Leu enkephalin sequence (H_2N -tyrosine, glycine, glycine, 5 phenylalanine, leucine- CO_2H , otherwise referred to herein as YGGFL (SEQ. ID NO:1)) was attached via its carboxy end to the exposed amino groups on the surface of the slide. The peptide was added in DMF solution with the BOP/HOBT/DIEA coupling reagents and recirculated through the flow cell for 2 hours at room temperature.

A first antibody, known as the Herz antibody, was applied to the surface of the slide for 45 minutes at 2 $\mu\text{g}/\text{ml}$ in a supercocktail (containing 1% BSA and 1% ovalbumin also in this case). A second antibody, goat anti-mouse fluorescein 15 conjugate, was then added at 2 $\mu\text{g}/\text{ml}$ in the supercocktail buffer, and allowed to incubate for 2 hours.

The results of this experiment are obtained by taking a fluorescence scan obtained using a fluorescence detection system. Again, FIG. 10 illustrates fluorescence intensity as a function of position. The fluorescence scale is shown on the right. This image was taken at 10 μm steps. This figure indicates that not only can deprotection be carried out in a well defined pattern, but also that (1) the method provides for successful coupling of peptides to the surface of the substrate, (2) the surface of a bound peptide is available for binding with an antibody, and (3) that the detection apparatus capabilities are sufficient to detect binding of a receptor.

2. EXAMPLE

FIG. 11 is a flow chart illustrating another example of the invention. Carboxy-activated NVOC-leucine was allowed to react with an aminated substrate. The carboxy activated HOBT ester of leucine and other amino acids used in this synthesis was formed by mixing 0.25 mmol of the NVOC amino protected amino acid with 37 mg HOBT (1-hydroxybenzotriazole), 111 mg BOP (benzotriazolyl-n-oxy-tris (dimethylamino)phosphoniumhexafluorophosphate) and 86 μl DIEA (diisopropylethylamine) in 2.5 ml DMF. The NVOC protecting group was removed by uniform illumination. Carboxy-activated NVOC-phenylalanine was coupled to the exposed amino groups for 2 hours at room temperature, and then washed with DMF and methylene chloride. Two unmasked cycles of photo-deprotection and coupling with carboxy-activated NVOC-glycine were carried out. The surface was then illuminated through a chrome on glass 50 μm checkerboard pattern mask. Carboxy-activated Na-tBOC-O-tButyl-L-tyrosine was then added. The entire surface was uniformly illuminated to photolyze the remaining NVOC groups. Finally, carboxy-activated NVOC-L-proline was added, the NVOC group was removed by illumination, and the t-BOC and t-butyl protecting groups were removed with TFA. After removal of the protecting groups, the surface consisted of a 50 μm checkerboard array of Tyr-Gly-Gly-Phe-Leu (YGGFL) and Pro-Gly-Gly-Phe-Leu (PGGFL). See also SEQ ID NO:1 and SEQ ID NO:2.

The array of pentapeptides was probed with a mouse monoclonal antibody directed against β -endorphin. This antibody (called 3E7) is known to bind YGGFL and YGGFM (see also SEQ ID NO:1 and SEQ ID NO:21) with nanomolar affinity and is discussed in Meo et al., *Proc. Natl. Acad. Sci. USA* (1983) 80:4084, which is incorporated by reference herein for all purposes. This antibody requires the amino terminal tyrosine for high affinity binding. The array of peptides formed as described in FIG. 11 was incubated

26

with a 2 $\mu\text{g}/\text{ml}$ mouse monoclonal antibody (3E7) known to recognize YGGFL. See also SEQ ID NO:1. 3E7 does not bind PGGFL. See also SEQ ID NO:2. A second incubation with fluoresceinated goat anti-mouse antibody labeled the regions that bound 3E7. The surface was scanned with an epi-fluorescence microscope. As shown in FIG. 12, results showed alternating bright and dark 50 μm squares indicating that YGGFL (SEQ ID NO:1) and PGGFL (SEQ ID NO:2) were synthesized in a geometric array determined by the mask. A high contrast ($>12:1$ intensity ratio) fluorescence checkerboard image shows that (a) YGGFL (SEQ ID NO:1) and PGGFL (SEQ ID NO:2) were synthesized in alternate 50 μm squares, (b) YGGFL (SEQ ID NO:1) attached to the surface is accessible for binding to antibody 3E7, and (c) antibody 3E7 does not bind to PGGFL (SEQ ID NO:2)

A three-dimensional representation of the fluorescence intensity data in a 2 square by 4 square rectangular portion of the checkerboard was also produced. It showed that the border between synthesis sites is sharp. The height of each spike in this display is linearly proportional to the integrated fluorescence intensity in a 2.5 μm pixel. The transition between PGGFL and YGGFL occurs within two spikes (5 μm). There is little variation in the fluorescence intensity of different YGGFL squares. The mean intensity of sixteen YGGFL synthesis sites was 2.03×10^5 counts and the standard deviation was 9.6×10^3 counts.

3. EXAMPLE

Monomer-by-Monomer Formation of YGGFL and Subsequent Exposure to Labeled Antibody

Monomer-by-monomer synthesis of YGGFL and GGFL in alternate squares was performed on a slide in a checkerboard pattern and the resulting slide was exposed to the Herz antibody. This experiment and the results thereof are illustrated in FIGS. 13A, 13B, 13C, and 13D.

In FIG. 13A, a slide is shown which is derivatized with t-BOC-aminopropyl-triethoxysilane. The slide was treated with TFA to remove the t-BOC protecting group. t-BOC-6-aminocaproic acid was then coupled onto the aminopropyl groups. The aminocaproic acid serves as a spacer between the aminopropyl group and the peptide to be synthesized. The amino end of the spacer was deprotected and coupled to NVOC-Leucine. The entire slide was then illuminated with 12 mW of 325 nm broadband illumination. The slide was then coupled with NVOC-Phenylalanine and washed. The entire slide was again illuminated, then coupled to NVOC-Glycine and washed. The slide was again illuminated and coupled to NVOC-Glycine to form the sequence shown in the last portion of FIG. 13A.

As shown in FIG. 13B, alternating regions of the slide were then illuminated using a projection print with a 500 \times 500 μm checkerboard mask; thus, the amino group of Glycine was exposed only in the lighted areas. When the next coupling chemistry step was carried out, NVOC-Tyrosine was added, and it coupled only at those spots which had received illumination. The entire slide was then illuminated to remove all the NVOC groups, leaving a checkerboard of YGGFL in the lighted areas and in the other areas, GGFL (SEQ. ID NO:15). The Herz antibody (which recognizes the YGGFL, but not GGFL) was then added, followed by goat anti-mouse fluorescein conjugate.

The resulting fluorescence scan is shown in FIG. 13C, and the scale for the fluorescence intensity is again given on the right. Dark areas contain the tetrapeptide GGFL, which is not recognized by the Herz antibody (and thus there is no binding of the goat anti-mouse antibody with fluorescein conjugate), and in the lightly shaded areas, YGGFL is

present. The YGGFL pentapeptide is recognized by the Herz antibody and, therefore, there is antibody in the lighted regions for the florescein-conjugated goat anti-mouse to recognize.

Similar patterns are shown for a 50 μm mask used in direct contact ("proximity print") with the substrate in FIG. 13D. Note that the pattern is more distinct and the corners of the checkerboard pattern are touching when the mask is placed in direct contact with the substrate (which reflects the increase in resolution using this technique).

4. EXAMPLE

Monomer-by-Monomer Synthesis of YGGFL and YPGGFL

In order to further demonstrate the operability of the invention, a 50 μm checkerboard pattern of alternating YGGFL and YPGGFL (SEQ. ID NO:3) was synthesized on a substrate using techniques like those set forth above. The resulting fluorescence plot is provided in FIG. 14. Again, it is seen that the antibody is clearly able to recognize the YGGFL sequence and does not bind significantly at the YPGGFL regions.

5. EXAMPLE

Synthesis of an Array of Sixteen Different Amino Acid Sequences and Estimation of Relative Binding Affinity to Herz Antibody

Using techniques similar to those set forth above, an array of 16 different amino acid sequences (replicated four times) was synthesized on each of two glass substrates. The sequences were synthesized by attaching the sequence NVOC-GFL across the entire surface of the slides. Using a series of masks, two layers of amino acids were then selectively applied to the substrate. Each region had dimensions of 0.25 cmx0.0625 cm. The first slide contained amino acid sequences containing only L amino acids while the second slide contained selected D amino acids. FIGS. 15A and 15B illustrate a map of the various regions on the first and second slides, respectively. The patterns shown in FIGS. 15A and 15B were duplicated four times on each slide. The slides were then exposed to the Herz antibody and fluorescein-labeled goat anti-mouse.

FIG. 16 is a fluorescence plot of the first slide, which contained only L amino acids. Light shading indicates strong binding (149,000 counts or more) while black indicates little or no binding of the Herz antibody (20,000 counts or less). The bottom right-hand portion of the slide appears "cut-off" because the slide was broken during processing. The sequence YGGFL is clearly most strongly recognized. The sequences YAGFL and YSGFL also exhibit strong recognition of the antibody. By contrast, most of the remaining sequences show little or no binding. The four duplicate portions of the slide are extremely consistent in the amount of binding shown therein.

FIG. 17 is a fluorescence plot of the second slide. Again, strongest binding is exhibited by the YGGFL sequence. Significant binding is also detected to YaGFL (SEQ. ID NO:22), YsGFL (SEQ. ID NO:23), and YpGFL (SEQ. ID NO:24). The remaining sequences show less binding with the antibody. Note the low binding efficiency of the sequence yGGFL.

Table 2 lists the various sequences tested in order of relative fluorescence, which provides information regarding relative binding affinity. In the table, lower case letters represent D-amino acids.

TABLE 2

Apparent Binding to Herz Ab	
L-a.a. Set	D-a.a. Set
YGGFL	YGGFL
YAGFL	YaGFL
YSGFL	YsGFL
LGGFL (SEQ. ID NO:25)	YpGFL
FGGFL (SEQ. ID NO:26)	fGGFL
YPGFL	yGGFL
LAGFL (SEQ. ID NO:27)	faGFL
FAGFL (SEQ. ID NO:28)	wGGFL
WGGFL (SEQ. ID NO:29)	yaGFL
	fpGFL
	waGFL

6. EXAMPLE

A 4096 compound experiment was conducted similarly to detailed Example 8. The building blocks used were: Y, G, P, A, F, W, G, F, M, Q, L, and S. Since G and F are repeated there are 4072 peptides. This chip was stained with Herz 3E7 IgG and then FITC-labelled goat anti-mouse IgG. A substrate used in this experiment was 700 μm thick. The results are shown in FIG. 18.

7. EXAMPLE

In order to generate 65,536 different compounds (including one null compound) in a minimum number of chemical steps, a sixteen step binary masking strategy was used. The building blocks chosen were (from amino to carboxy): r, R, H, Q, P, F, Homophenylalanine, N, Ornithine, A, V, v, T, S, and G (lower case letters represent D-amino acids). This experiment shows that unnatural amino acids can be used as building blocks.

Once the masking strategy and the building blocks were chosen, the amino acids were weighed into cartridges obtained from ABI (Applied Biosystems Inc., Foster City, Calif.). For this experiment the flow cell volume was 0.5 ml, and thus 15 mg of HOBt and 44 mg of BOP were used. During synthesis the amino acid was dissolved in 1.5 ml of solvent so that the flow cell was full of amino acid solution during coupling. The final amino acid concentration was 60 mM.

Next a process file was generated on the program "PS" (copy provided in Appendix 3). This was done by hitting F1 (if an IBM computer was being used to run the program) to "initialize masking sequence." "Binary process minimum movement" was chosen and then the program asked for input the building blocks in order of C terminus to N terminus. The first building block that was input that goes onto the chip first, in this case S. The program allows for input of the names in either one letter or three letter codes. Using a binary process with minimum movement, one does not have to select the mask that will be used, as the program will select the mask. For a sixteen-step binary synthesis the following masks are used in the order given: mask A offset 0, mask A offset 1, mask B offset 0, mask B offset 1, mask C offset 0, mask C offset 1, mask D offset 0, mask D offset 1, mask E offset 0, mask E offset 1, mask F offset 0, mask F offset 1, mask G offset 0, mask G offset 1, mask H offset 0, and mask H offset 1. The masking sequence was then saved to disk so that it could be used during data workup.

The exposure lamp was turned on and the shutter timer set for 11 minutes. The lamp power was set to about 12 mW/cm² at 365 nm. The lamp was warmed up for about an

US 6,420,169 B1

29

hour before the experiment began. Meanwhile, the mask was cleaned with methanol and the mirror was aligned.

Next, the ABI peptide synthesizer was set up. Of course, other peptide synthesizers can be used, e.g., one commercially available from Milligen, Inc. All of the reagent bottles were filled and an empty cartridge was placed under the injector. The amino acid cartridges were then loaded from the N-terminus on the left to the C-terminus on the right. The synthesizer modules to be used were then entered through synthesizer interface. In this case the following sequence was used: HEBDCDFDCCD. This sequence is repeated sixteen times for a sixteen step synthesis.

H is the photolysis module. The slide was rinsed with p-dioxane and then rinsed twice in 5 mM H₂SO₄ in p-dioxane. The flow cell was then filled with 5 mM H₂SO₄ in p-dioxane for photolysis. A relay was sent to the personal computer telling it to open the lamp shutter. The shutter opens and remains open for 11 minutes. During this time a 5% DIEA in DMF solution was made in the peptide synthesizer's activator vessel. Module E begins before the photolysis is completed. The first part of module E starts the dissolution of the amino acid in the cartridge. A 7% DIEA/DMF solution was delivered to the cartridge and the solution was mixed by Argon bubbling. The cartridge was mixed for about six minutes. Next the cartridge solution was further diluted with anhydrous DMF and mixed some more. By this point the photolysis was completed.

Module B, the substrate activation module, first rinsed the slide with p-dioxane six times. Then it rinsed the slide with DMF. Next, an aliquot of the 5% DIEA/DMF solution made in the activator vessel was moved to the flow cell where it sat for 100 seconds. This step was repeated six times and the flow cell was drained.

Module D washed the slide with DMF and washed the slide with dichloromethane followed by ethanol. Hence, the slide was washed with DMF, dichloromethane and ethanol, and finally DMF again.

Next chemical coupling occurred using module F. The amino acid solution was taken from the cartridge and put into the flow cell where it sat for 1.5 hours. After coupling was completed, the slide was then washed using modules D, C, C, and D. This is the end of one synthesis cycle. As mentioned above, this sequence of modules was repeated fifteen more times for a sixteen step synthesis.

In use, the flow cell was set up and a substrate was chosen. In this case a 160 μ m thick slide derivatized with an NVOC-6-aminocaproic acid linker was used. The slide was placed on the flow cell and the vacuum was turned on. The flow cell transfer lines were attached to the synthesizer. The slide and flow cell were checked for leaks using a methylene chloride wash. This also served to rinse the slide. Next, the outer surface of the slide was cleaned with methanol. The flow cell was then attached to the synthesis mount and placed flush against the mask.

The synthesis then began. "Begin synthesis" was pressed on "PS," (copy provided in Appendix 3) with the exposure time set to 660 seconds. After the mask moved to its first position, the ABI was started. This synthesis took about 48 hours, since each cycle was three hours.

Once the synthesis was completed, the slide underwent a final photolysis to remove all of the NVOC groups on the slide. Modules HIBDCD were used, with I a "wait" step. Since it is desirable to photolyze the entire slide, the flow cell with the attached substrate was taken off of the synthesis mount and physically placed under the lamp. After the photolysis was complete, it was put back so that the remain-

30

ing modules would go smoothly. The flow cell was in a vertical position to ensure total coverage of the substrate with solutions.

The amino groups were capped by final photolysis with acetic anhydride. This process is called acetylation.

After final photolysis and capping, the side chain protecting groups on the amino acids were removed. The slide was taken off of the flow cell and treated with a trifluoroacetic acid solution containing phenol, thioanisole, and ethanedithiol as scavengers. After side-group deprotection the slide was neutralized in a 5% DIEA/methylene chloride solution twice for five minutes each. The slide was then rinsed with methylene chloride, DMF, and ethanol.

Next, the slide was incubated with 3 ml of anti-dynorphin B antibody (8 micrograms/ml) and 1% BSA in PBS (containing 0.08% Tween 20™) for two hours. After rinsing twice with PBS, the slide was stained with FITC-labelled goat anti-mouse antibody (10 micrograms/ml) in 1% BSA/PBS for 1.5 hours. After the second staining, the slide was rinsed twice with PBS/Tween 20™ and once with deionized water.

Next the slide was scanned using the fluorescence detection system. Scanning parameters depend on the type of image being scanned. In this case since each synthesis site is only 50 μ m, the slide was scanned very slowly with a small increment size. Typical parameters were 3000x3000 @5 μ m steps, 5 μ m/ms, 220 μ m/ms².

Once the fluorescence images were obtained, the data file was converted to a tiff file and analyzed using a program called "avi" (attached as Appendix 1) which has a module that integrates each synthesis site. It then made a file containing fluorescence versus location information. Then another program called "pepsrch" (attached as Appendix 2) was used to combine the fluorescence information with compound identity. In this experiment the largest peptide synthesis was 16 amino acids in length, and 65,536 (including the null) peptides (including monomers) were synthesized. The results are shown in FIG. 19.

C. Fluorescence Energy-Transfer Substrate Assays

A different application of the present invention tests for catalytic cleavage of various polymer sequences by an enzyme or other catalyst. For example, aspartyl proteases such as renin, HIV proteases, elastase, collagenase and some cathepsins can be tested against an array of peptides. According to this aspect of the invention, a variety of peptide sequences are synthesized on a solid substrate by the protection-deprotection strategy outlined above. The resulting array is probed with an enzyme which might cleave one or more peptide elements of the array resulting in a detectable chain.

In one embodiment, the peptides to be tested have a fluorescence donor group such as 1-aminobenzoic acid (anthranilic acid or ABZ) or aminomethylcoumarin (AMC) located at one position on the peptide and a fluorescence quencher group such as lucifer yellow, methyl red or nitrobenzo-2-oxo-1,3-diazole (NBD) at a different position near the distal end of the peptide. Note, that some "donor" groups can also serve as "quencher" groups, depending on the relative excitation and emission frequencies of the particular pair selected. The intramolecular resonance energy transfer from the fluorescence donor molecule to the quencher will quench the fluorescence of the donor molecule. Upon cleavage, however, the quencher is separated from the donor group, leaving behind a fluorescent fragment. Plus, a scan of the surface with an epifluorescence microscope for example, will show bright regions where the

US 6,420,169 B1

31

peptide has been cleaved. As an example, FIG. 20A shows a tripeptide having a donor-quencher pair on a substrate. The fluorescence donor molecule, 1-aminobenzoic acid (ABZ), is coupled to the ϵ -amino group of lysine (Lys) on the P' side of the substrate. The donor molecule could, of course, be attached to the α -amine group. A fluorescence quencher, NBD caproic acid is coupled to the P side of the substrate molecule. Upon cleavage by a protease as shown in FIG. 20B, the quencher is released leaving the fluorescent type fragment still bound to the solid substrate for detection.

FIG. 21 demonstrates the fluorescence donor-quencher resonance energy transfer assay for two quenchers. In FIG. 21A the bright outer circle is produced by fluorescence from benzoyloxycarbonyl protected 1-aminobenzoic acid linked to a slide through lysine and a linker. The dark inner circle shows the quenching effect of methyl red. FIG. 21B shows a similar result, with the dark inner circle resulting from the quenching effect of NBD-caproic acid.

III. Synthesis

A. Reactor System

FIG. 22a schematically illustrates a preferred embodiment of a reactor system 100 for synthesizing polymers on the prepared substrate in accordance with one aspect of the invention. The reactor system includes a body 102 with a cavity 104 on a surface thereof. In preferred embodiments the cavity 104 is between about 50 and 1000 μm deep with a depth of about 500 μm preferred.

The bottom of the cavity is preferably provided with an array of ridges 106 which extend both into the plane of the Figure and parallel to the plane of the Figure. The ridges are preferably about 50 to 200 μm deep and spaced at about 2 to 3 mm. The purpose of the ridges is to generate turbulent flow for better mixing. The bottom surface of the cavity is preferably light absorbing so as to prevent reflection of impinging light.

A substrate 112 is mounted above the cavity 104. The substrate is provided along its bottom surface 114 with a photoremovable protecting group such as NVOC with or without an intervening linker molecule. The substrate is preferably transparent to a wide spectrum of light, but in some embodiments is transparent only at a wavelength at which the protecting group may be removed (such as UV in the case of NVOC). The substrate in some embodiments is a conventional microscope glass slide or cover slip. The substrate is preferably as thin as possible, while still providing adequate physical support. Preferably, the substrate is less than about 1 mm thick, more preferably less than 0.5 mm thick, more preferably less than 0.1 mm thick, and can be less than 0.05 mm thick. In alternative preferred embodiments, the substrate is quartz, silicon, or other compounds such as a silicon nitride.

The substrate and the body serve to seal the cavity except for an inlet port 108 and an outlet port 110. The body and the substrate may be mated for sealing in some embodiments with one or more gaskets. According to one embodiment, the body is provided with two concentric gaskets and the intervening space is held at vacuum to ensure mating of the substrate to the gaskets.

Fluid is pumped through the inlet port into the cavity by way of a pump 116 which may be, for example, a model no. B-120-S made by Eldex Laboratories, from fluid supply 118. Selected fluids are circulated into the cavity by the pump, through the cavity, and out the outlet for recirculation or disposal. The reactor may be subjected to ultrasonic radiation and/or heated to aid in agitation in some embodiments.

Above the substrate 112, a lens 120 is provided which may be, for example, a 2" 100 mm focal length fused silica

32

lens. For the sake of a compact system, a reflective mirror 122 may be provided for directing light from a light source 124 onto the substrate. Light source 124 may be, for example, a Hg(Xe) light source manufactured by Oriel and having model no. 66024. A second lens 126 may be provided for the purpose of projecting a mask image onto the substrate in combination with lens 620. This form of lithography is referred to herein as projection printing. As will be apparent from this disclosure, proximity printing and the like may also be used according to some embodiments.

Light from the light source is permitted to reach only selected locations on the substrate as a result of mask 128. Mask 128 may be, for example, a glass slide having etched chrome thereon. The mask 128 in one embodiment is provided with a grid of transparent locations and opaque locations. Such masks may be manufactured by, for example, Photo Sciences, Inc. Light passes freely through the transparent regions of the mask, but is reflected from or absorbed by other regions. Therefore, only selected regions of the substrate are exposed to light.

As discussed above, light valves (LCD's) may be used as an alternative to conventional masks to selectively expose regions of the substrate. Alternatively, fiber optic faceplates such as those available from Schott Glass, Inc. may be used for the purpose of contrast enhancement of the mask or as the sole means of restricting the region to which light is applied. Such faceplates would be placed directly above or on the substrate in the reactor shown in FIG. 22A. In still further embodiments, flys-eye lenses, tapered fiber optic faceplates, or the like, may be used for contrast enhancement.

In order to provide for illumination of regions smaller than a wavelength of light, more elaborate techniques may be utilized. For example, according to one preferred embodiment, light is directed at the substrate by way of molecular microcrystals on the tip of, for example, micropipettes. Such devices are disclosed in Lieberman et al., "A Light Source Smaller Than the Optical Wavelength," *Science* (1990) 247:59-61, which is incorporated herein by reference for all purposes.

In operation, the substrate is placed on the cavity and sealed thereto. All operations in the process of preparing the substrate are carried out in a room lit primarily or entirely by light of a wavelength outside of the light range at which the protecting group is removed. For example, in the case of NVOC, the room should be lit with a conventional dark room light which provides little or no UV light. All operations are preferably conducted at about room temperature.

A first, deprotection fluid (without a monomer) is circulated through the cavity. The solution preferably is of 5 mM sulfuric acid in dioxane solution which serves to keep exposed amino groups protonated and decreases their reactivity with photolysis by-products. Absorptive materials such as N,N-diethylamino 2,4-dinitrobenzene, for example, may be included in the deprotection fluid which serves to absorb light and prevent reflection and unwanted photolysis.

The slide is, thereafter, positioned in a light ray path from the mask such that first locations on the substrate are illuminated and, therefore, deprotected. In preferred embodiments the substrate is illuminated for between about 1 and 15 minutes with a preferred illumination time of about 10 minutes at 10-20 mW/cm² with 365 nm light. The slides are neutralized (i.e., brought to a pH of about 7) after photolysis with, for example, a solution of diisopropylethylamine (DIEA) in methylene chloride for about 5 minutes.

The first monomer is then placed at the first locations on the substrate. After irradiation, the slide is removed, treated

US 6,420,169 B1

33

in bulk, and then reinstalled in the flow cell. Alternatively, a fluid containing the first monomer, preferably also protected by a protective group, is circulated through the cavity by way of pump **116aF**. If, for example, it is desired to attach the amino acid Y to the substrate at the first locations, the amino acid Y (bearing a protecting group on its α -nitrogen), along with reagents used to render the monomer reactive, and/or a carrier, is circulated from a storage container **118aF**, through the pump, through the cavity, and back to the inlet of the pump.

The monomer carrier solution is, in a preferred embodiment, formed by mixing of a first solution (referred to herein as solution "A") and a second solution (referred to herein as solution "B"). Table 3 provides an illustration of a mixture which may be used for solution A.

TABLE 3

Representative Monomer Carrier Solution "A"
0.25 mMoles NVOC amino protected amino acid
37 mg HOBt (1-Hydroxybenzotriazole)
250 μ l DMF (Dimethylformamide)
86 μ l DIEA (Diisopropylethylamine)

The composition of solution B is illustrated in Table 4. Solutions A and B are mixed and allowed to react at room temperature for about 8 minutes, then diluted with 2 ml of DMF, and 500 μ l are applied to the surface of the slide or the solution is circulated through the reactor system and allowed to react for about 2 hours at room temperature. The slide is then washed with DMF, methylene chloride and ethanol.

TABLE 4

Representative Monomer Carrier Solution "B"
250 μ l DMF
111 mg BOP (Benzotriazolyl-n-oxy-tris(dimethylamino) phosphoniumhexafluorophosphate)

As the solution containing the monomer to be attached is circulated through the cavity, the amino acid or other monomer will react at its carboxy terminus with amino groups on the regions of the substrate which have been deprotected. Of course, while the invention is illustrated by way of circulation of the monomer through the cavity, the invention could be practiced by way of removing the slide from the reactor and submersing it in an appropriate monomer solution.

After addition of the first monomer, the solution containing the first amino acid is then purged from the system. After circulation of a sufficient amount of the DMF/dichloromethane such that removal of the amino acid can be assured (e.g., about 50 \times times the volume of the cavity and carrier lines), the mask or substrate is repositioned, or a new mask is utilized such that second regions on the substrate will be exposed to light and the light **124** is engaged for a second exposure. This will deprotect second regions on the substrate and the process is repeated until the desired polymer sequences have been synthesized.

The entire derivatized substrate is then exposed to a receptor of interest, preferably labeled with, for example, a fluorescent marker, by circulation of a solution or suspension of the receptor through the cavity or by contacting the surface of the slide in bulk. The receptor will preferentially bind to certain regions of the substrate which contain complementary sequences.

Antibodies are typically suspended in what is commonly referred to as "supercocktail," which may be, for example, a solution of about 1% BSA (bovine serum albumin), 0.05% Tween 20TM in PBS (phosphate buffered saline) buffer. The antibodies are diluted into the supercocktail buffer to a final concentration of, for example, about 0.1 to 4 μ g/ml.

FIG. 22B illustrates an alternative preferred embodiment of the reactor shown in FIG. 22A. According to this

34

embodiment, the mask **128** is placed directly in contact with the substrate. Preferably, the etched portion of the mask is placed face down so as to reduce the effects of light dispersion. According to this embodiment, the imaging lenses **120** and **126** are not necessary because the mask is brought into close proximity with the substrate.

FIG. 23 schematically illustrates a particularly preferred device used to synthesize diverse polymer sequences on a substrate. The device includes an automated peptide synthesizer **401**. The automated peptide synthesizer is a device which flows selected reagents through a flow cell **402** under the direction of a computer **404**. In a preferred embodiment the synthesizer is an ABI Peptide Synthesizer, model no. 431A. The computer may be selected from a wide variety of computers or discrete logic including for, example, an IBM PC-AT or similar computer linked with appropriate internal control systems in the peptide synthesizer. The PC is provided with signals from the ABI computer indicative of, for example, the beginning of a photolysis cycle. One can also modify the synthesizer with a board that links the contacts of relays in the computer in parallel with the switches to the keyboard of the control panel of the synthesizer to eliminate some of the keystrokes that would otherwise be required to operate the synthesizer.

Substrate **406** is mounted on the flow cell, forming a cavity between the substrate and the flow cell. Selected reagents flow through this cavity from the peptide synthesizer at selected times, forming an array of peptides on the face of the substrate in the cavity. Mounted above the substrate, and preferably in contact with the substrate is a mask **408**. Mask **408** is transparent in selected regions to a selected wavelength of light and is opaque in other regions to the selected wavelength of light. The mask is illuminated with a light source **410** such as a UV light source. In one specific embodiment the light source **410** is a model no. 82420 made by Oriel. The mask is held and translated by an x-y translation stage **412** such as a translation stage made by Newport corp. The computer coordinates action of the peptide synthesizer, translation stage, and light source. Of course, the invention may be used in some embodiments with translation of the substrate instead of the mask.

In operation, the substrate is mounted on the flow cell. The substrate, with its surface protected by a suitable photo removable protecting group, is exposed to light at selected locations by positioning the mask and directing light from a light source, through the mask, onto the substrate for a desired period of time (such as, for example, 1 sec to 60 min in the case of peptide synthesis). A selected peptide or other monomer/polymer is pumped through the reactor cavity by the peptide synthesizer for binding at the selected locations on the substrate. After a selected reaction time (such as about 1 sec to 300 min in the case of peptide reactions) the monomer is washed from the system, the mask is appropriately repositioned or replaced, and the cycle is repeated. In most embodiments of the invention, reactions may be conducted at or near ambient temperature. Agitation can be used to mix the reaction contents.

FIGS. 24A and 24B are flow charts of the software used in operation of the reactor system. At step **502** the peptide synthesis software (PS, attached as appendix 3) is initialized. At step **504** the system calibrates positioners on the x-y translation stage and begins a main loop. At step **506** the system determines which, if any, of the function keys on the computer have been pressed. If F1 has been pressed, the system prompts the user for input of a desired synthesis process. If the user enters F2, the system allows a user to edit a file for a synthesis process at step **510**. If the user enters

US 6,420,169 B1

35

F3 the system loads a process from a disk at step 512. If the user enters F4 the system saves an entered or edited process to disk at step 514. If the user selects F5 the current process is displayed at step 516 while selection of F6 starts the main portion of the program, i.e., the actual synthesis according to the selected process. If the user selects F7 the system displays the location of the synthesized peptides, while pressing F10 returns the user to the disk operating system.

FIG. 24B illustrates the synthesis step 518 in greater detail. The main loop of the program is started in which the system first moves the mask to a next position at step 526. During the main loop of the program, necessary chemicals flow through the reaction cell under the direction of the on-board computer in the peptide synthesizer. At step 528 the system then waits for an exposure command and, upon receipt of the exposure command exposes the substrate for a desired time at step 530. When an acknowledgement of complete exposure is received at step 532 the system determines if the process is complete at step 534 and, if so, waits for additional keyboard input at step 536 and, thereafter, exits the perform synthesis process.

A computer program ("PS") used for operation of the system described above is written in Turbo C (Borland Int'l) and has been implemented in an IBM compatible system. The motor control software is adapted from software produced by Newport Corporation. It will be recognized that a large variety of programming languages could be utilized without departing from the scope of the invention herein. Certain calls are made to a graphics program in "Programmer Guide to PC and PS2 Video Systems" (Wilton, Microsoft Press, 1987), which is incorporated herein by reference for all purposes.

Alignment of the mask is achieved by one of two methods in preferred embodiments. In a first embodiment the system relies upon relative alignment of the various components, which is normally acceptable since x-y-z translation stages are capable of sufficient accuracy for the purposes herein. In alternative embodiments, alignment marks on the substrate are coupled to a CCD device for appropriate alignment.

According to some embodiments, pure reagents are not added at each step, or complete photolysis of the protecting groups is not provided at each step. According to these embodiments, multiple products will be formed in each synthesis site. For example, if the monomers A and B are mixed during a synthesis step, A and B will bind to deprotected regions, roughly in proportion to their concentration in solution. Hence, a mixture of compounds will be formed in a synthesis region. A substrate formed with mixtures of compounds in various synthesis regions may be used to perform, for example, an initial screening of a large number of compounds, after which a smaller number of compounds in regions which exhibit high binding affinity are further screened. Similar results may be obtained by only partially photolyzing a region, adding a first monomer, re-photolyzing the same region, and exposing the region to a second monomer.

B. Combinatorial Synthesis Strategy

In preferred embodiments an ordered sequence of masks is utilized. In some embodiments it is possible to use as few as a single mask to synthesize all of the possible polymers of a given monomer set.

If, for example, it is desired to synthesize all 16 dinucleotides from four bases, a 1 cm square synthesis region is divided conceptually into 16 boxes, each 0.25 cm wide. Denote the four monomer units by A, B, C, and D. The first reactions are carried out in four vertical columns, each 0.25 cm wide. The first mask exposes the left-most column of

36

boxes, where A is coupled. The second mask exposes the next column, where B is coupled; followed by a third mask, for the C column; and a final mask that exposes the right-most column, for D. The first, second, third, and fourth masks may be a single "all-purpose" mask translated to different locations. Such an "all-purpose" mask can be useful in any synthesis strategy, whether binary or not.

The process is repeated in the horizontal direction for the second unit of the dimer. This time, the masks allow exposure of horizontal rows, again 0.25 cm wide. A, B, C, and D are sequentially coupled using masks that expose horizontal fourths of the reaction area. The resulting substrate contains all 16 dinucleotides of four bases.

The eight masks used to synthesize the dinucleotides or other dimers are related to one another by translation or rotation. In fact, one mask can be used in all eight steps if it is suitably rotated and translated. For example, in the example above, a mask with a single transparent region could be sequentially used to expose each of the vertical columns, translated 90°, and then sequentially used to allow exposure of the horizontal rows.

Tables 5 and 6 provide a simple computer program in Quick Basic for planning a masking program and a sample output, respectively, for the synthesis of a polymer chain of three monomers ("residues") having three different monomers in the first level, four different monomers in the second level, and five different monomers in the third level in a striped pattern. The output of the program is the number of cells, the number of "stripes" (light regions) on each mask, and the amount of translation required for each exposure of the mask.

TABLE 5

Mask Strategy Program

```

DEFINT A-Z
DIM b(20), w(20), l(500)
F$ = "LPT1:"
OPEN F$ FOR OUTPUT AS #1
jmax = 3 'Number of residues
b(1) = 3: b(2) = 4: b(3) = 5 'Number of building blocks for res 1,2,3
g = 1: lmax(1) = 1
FOR j = 1 TO jmax: g = g * b(j): NEXT j
w(0) = 0: w(1) = g / b(1)
PRINT #1, "MASK2.BAS ", DATE$, TIME$: PRINT #1,
PRINT #1, USING "Number of residues=##": jmax
FOR j = 1 TO jmax
PRINT #1, USING "    Residue ##    ## building blocks": j; b(j)
NEXT j
PRINT #1,
PRINT #1, USING "Number of cells = ####": g: PRINT #1,
FOR j = 2 TO jmax
lmax(j) = lmax(j - 1) * b(j - 1)
w(j) = w(j - 1) / b(j)
NEXT j
FOR j = 1 TO jmax
PRINT #1, USING "Mask for residue ##": j: PRINT #1,
PRINT #1, USING "    Number of stripes = ####": lmax(j)
PRINT #1, USING "    Width of each stripe = ####": w(j)
FOR l = 1 TO lmax(j)
a = 1 + (l - 1) * w(j - 1)
ae = a + w(j) - 1
PRINT #1, USING "    Stripe ## begins at location #### and
ends at ####": l; a; ae
NEXT l
*PRINT #1,
PRINT #1, USING "    For each of ## building blocks,
translate mask by ##
cell(s)": b(j); w(j),
PRINT #1, : PRINT #1, : PRINT #1,
NEXT j

```


US 6,420,169 B1

37

© Copyright 1990, Affymax Technologies N.V.

TABLE 6

Masking Strategy Output	
Number of residues = 3	
Residue 1	3 building blocks
Residue 2	4 building blocks
Residue 3	5 building blocks
Number of cells = 60	
Mask for residue 1	
Number of stripes = 1	
Width of each stripe = 20	
Stripe 1 begins at location 1 and ends at 20	
For each of 3 building blocks, translate mask by 20 cell(s)	
Mask for residue 2	
Number of stripes = 3	
Width of each stripe = 5	
Stripe 1 begins at location 1 and ends at 5	
Stripe 2 begins at location 21 and ends at 25	
Stripe 3 begins at location 41 and ends at 45	
For each of 4 building blocks, translate mask by 5 cell(s)	
Mask for residue 3	
Number of stripes = 12	
Width of each stripe = 1	
Stripe 1 begins at location 1 and ends at 1	
Stripe 2 begins at location 6 and ends at 6	
Stripe 3 begins at location 11 and ends at 11	
Stripe 4 begins at location 16 and ends at 16	
Stripe 5 begins at location 21 and ends at 21	
Stripe 6 begins at location 26 and ends at 26	
Stripe 7 begins at location 31 and ends at 31	
Stripe 8 begins at location 36 and ends at 36	
Stripe 9 begins at location 41 and ends at 41	
Stripe 10 begins at location 46 and ends at 46	
Stripe 11 begins at location 51 and ends at 51	
Stripe 12 begins at location 56 and ends at 56	
For each of 5 building blocks; translate mask by 1 cell(s)	

© Copyright 1990, Affymax Technologies N.V.

In preferred embodiments of the invention herein a binary synthesis strategy is utilized. The binary synthesis strategy is illustrated herein primarily with regard to a masking strategy, although it will be applicable to other polymer synthesis strategies such as the pin strategy, and the like.

In a binary synthesis strategy, the substrate is irradiated with a first mask, exposed to a first building block, irradiated with a second mask, exposed to a second building block, etc. Each combination of masked irradiation and exposure to a building block is referred to herein as a "cycle."

In a preferred binary masking strategy, the masks for each cycle allow illumination of half of a region of interest on the substrate and no illumination of the remaining half of the region of interest. By "half" it is intended herein not to mean exactly one-half the region of interest, but instead a large fraction of the region of interest such as from about 30 to 70 percent of the region of interest. It will be understood that the entire masking strategy need not take a binary form; instead non-binary cycles may be introduced as desired between binary cycles.

In preferred embodiments of the binary masking strategy, a given cycle illuminates only about half of the region which was illuminated in a previous cycle, while not illuminating the remaining half of the illuminated portion from the previous cycle. Conversely, in such preferred embodiments, a given cycle illuminates half of the region which was not illuminated in the previous cycle and does not illuminate half the region which was not illuminated in a previous cycle.

The synthesis strategy is most readily illustrated and handled in matrix notation. At each synthesis site, the determination of whether to add a given monomer is a binary process. Therefore, each product element P_j in a product

38

matrix P is given by the dot product of two vectors, a chemical reactant vector, (CRV) e.g., $CRV=[A,B,C,D]$, and a binary vector σ_j . Inspection of the products in the example below for a four-step synthesis, shows that in one four-step synthesis $\sigma_1=[1,0,1,0]$, $\sigma_2=[1,0,0,1]$, $\sigma_3=[0,1,1,0]$, and $\sigma_4=[0,1,0,1]$, where a 1 indicates illumination and a 0 indicates no illumination. Therefore, it becomes possible to build a "switch matrix" S from the column vectors σ_j ($j=1,k$ where k is the number of products).

$$S = \begin{matrix} & \sigma_1 & \sigma_2 & \sigma_3 & \sigma_4 \\ \begin{matrix} 1 \\ 10 \\ 15 \end{matrix} & \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \end{matrix}$$

The outcome P of a synthesis is simply $P=CS$, the product of the chemical reactant matrix and the switch matrix.

The switch matrix for an n -cycle synthesis yielding k products has n rows and k columns. An important attribute of S is that each row specifies a mask. A two-dimensional mask m_j for the j th chemical step of a synthesis is obtained directly from the j th row of S by placing the elements $s_{j1} \dots s_{jk}$ into, for example, a square format. The particular arrangement below provides a square format, although linear or other arrangements may be utilized.

$$S = \begin{matrix} & s_{11} & s_{12} & s_{13} & s_{14} \\ \begin{matrix} 30 \\ 40 \end{matrix} & \begin{bmatrix} s_{21} & s_{22} & s_{23} & s_{24} \\ s_{31} & s_{32} & s_{33} & s_{34} \\ s_{41} & s_{42} & s_{43} & s_{44} \end{bmatrix} & m_j = \begin{matrix} s_{j1} & s_{j2} \\ s_{j3} & s_{j4} \end{matrix} \end{matrix}$$

Of course, compounds formed in a light-activated synthesis can be positioned in any defined geometric array. A square or rectangular matrix is convenient but not required. The rows of the switch matrix may be transformed into any convenient array as long as equivalent transformations are used for each row.

For example, the masks in the four-step synthesis below are then denoted by:

$$m_1 = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \quad m_2 = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix} \quad m_3 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \quad m_4 = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$$

where 1 denotes illumination (activation) and 0 denotes no illumination.

The matrix representation is used to generate a desired set of products and product maps in preferred embodiments. Each compound is defined by the product of the chemical vector and a particular switch vector. Therefore, for each synthesis address, one simply saves the switch vector, assembles all of them into a switch matrix, and extracts each of the rows to form the masks.

In some cases, particular product distributions or a maximal number of products are desired. For example, for $CRV=[A,B,C,D]$, any switch vector (σ_j) consists of four bits. Sixteen four-bit vectors exist. Hence, a maximum of 16 different products can be made by sequential addition of the reagents $[A,B,C,D]$. These 16 column vectors can be assembled in 16! different ways to form a switch matrix. The order of the column vectors defines the masking patterns, and therefore, the spatial ordering of products but not their makeup. One ordering of these columns gives the following switch matrix (in which "null" (\emptyset) additions are included in

US 6,420,169 B1

39

brackets for the sake of completeness, although such null additions are elsewhere ignored herein):

σ_1	σ_{16}	CRV
1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0	A	5
[0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1]	\emptyset	
1 1 1 1 0 0 0 0 1 1 1 1 0 0 0 0	B	
$S =$ [0 0 0 0 1 1 1 1 0 0 0 0 1 1 1 1]	\emptyset	
1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0	C	10
[0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1]	\emptyset	
1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0	D	
[0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1]	\emptyset	

The columns of S according to this aspect of the invention are the binary representations of the numbers 15 to 0. The sixteen products of this binary synthesis are ABCD, ABC, ABD, AB, ACD, AC, AD, A, BCD, BC, BD, B, CD, C, D, and \emptyset (null). Also note that each of the switch vectors from the four-step synthesis masks above (and hence the synthesis products) are present in the four bit binary switch matrix. (See columns 6, 7, 10, and 11). Note that if the desired compounds comprise only dimers, then one could extract the switch vectors for compounds AB, AC, AD, BC, BD, and CD for this synthesis.

This synthesis procedure provides an easy way for mapping the completed products. The products in the various locations on the substrate are simply defined by the columns of the switch matrix (the first column indicating, for example, that the product ABCD will be present in the upper left-hand location of the substrate). Furthermore, if only selected desired products are to be made, the mask sequence can be derived by extracting the columns with the desired sequences. For example, to form the product set ABCD, ABD, ACD, AD, BCD, BD, CD, and D, the masks are formed by use of a switch matrix with only the 1st, 3rd, 5th, 7th, 9th, 11th, 13th, and 15th columns arranged into the switch matrix:

$S =$	1 1 1 1 0 0 0 0
	1 1 0 0 1 1 0 0
	1 0 1 0 1 0 1 0
	1 1 1 1 1 1 1 1

To form all of the polymers of length 4, the reactant matrix [ABCDABCDABCDABCD] is used. The switch matrix will be formed from a matrix of the binary numbers from 0 to 2^{16} arranged in columns. The columns having four monomers are then selected and arranged into a switch matrix. Therefore, it is seen that the binary switch matrix in general will provide a representation of all the products which can be made from an n-step synthesis, from which the desired products are then extracted.

The rows of the binary switch matrix will, in preferred embodiments, have the property that each masking step illuminates half of the synthesis area. Each masking step also factors the preceding masking step; that is, half of the region that was illuminated in the preceding step is again illuminated, whereas the other half is not. Half of the region that was not illuminated in the preceding step is also illuminated, whereas the other half is not. Thus, masking is recursive. The masks are constructed, as described previously, by extracting the elements of each row and placing them in a square array. For example, the four masks in S for a four-step synthesis are:

40

$m_1 =$	1 1 1 1	$m_2 =$	1 1 1 1	$m_3 =$	1 1 0 0
	1 1 1 1		0 0 0 0		1 1 0 0
	0 0 0 0		1 1 1 1		1 1 0 0
	0 0 0 0		0 0 0 0		1 1 0 0
	1 0 1 0				
$m_4 =$	1 0 1 0				
	1 0 1 0				
	1 0 1 0				

The recursive factoring of masks allows the products of a light-directed synthesis to be represented by a polynomial. (Some light activated syntheses can only be denoted by irreducible, i.e., prime polynomials.) For example, the polynomial corresponding to the top synthesis of FIG. 8a (discussed below) is

$$P=(A+B)(C+D)$$

A reaction polynomial may be expanded as though it were an algebraic expression, provided that the order of joining of reactants X_1 and X_2 is preserved ($X_1X_2 \neq X_2X_1$), i.e., the products are not commutative. The product then is $AC+AD+BC+BD$. The polynomial explicitly specifies the reactants and implicitly specifies the mask for each step. Each pair of parentheses demarcates a round of synthesis. The chemical reactants of a round (e.g., A and B) react at nonoverlapping sites and hence cannot combine with one another. The synthesis area is divided equally among the elements of a round (e.g., A is directed to one-half of the area and B to the other half). Hence, the masks for a round (e.g., the masks m_A and m_B) are orthogonal and form an orthonormal set. The polynomial notation also signifies that each element in a round is to be joined to each element of the next round (e.g., A with C, A with D, B with C, and B with D). This is accomplished by having m_C overlap m_A and m_B equally, and likewise for m_D . Because C and D are elements of a round, m_C and m_D are orthogonal to each other and form an orthonormal set.

The polynomial representation of the binary synthesis described above, in which 16 products are made from 4 reactants, is

$$P=(A+\emptyset)(B+\emptyset)(C+\emptyset)(D+\emptyset)$$

which gives ABCD, ABC, ABD, AB, ACD, AC, AD, A, BCD, BC, BD, B, CD, C, D, and \emptyset when expanded (with the rule that $\emptyset X=X$ and $X\emptyset=X$, and remembering that joining is ordered). In a binary synthesis, each round contains one reactant and one null (denoted by \emptyset). Half of the synthesis area receives the reactant and the other half receives nothing. Each mask overlaps every other mask equally.

Binary rounds and non-binary rounds can be interspersed as desired, as in

$$P=(A+\emptyset)(B)(C+D+\emptyset)(E+F+G)$$

The 18 compounds formed are ABCE, ABCE, ABCG, ABDE, ABDF, ABDG, ABE, ABF, ABG, BCE, BCF, BCG, BDE, BDF, BDG, BE, BF, and BG. The switch matrix S for this 7-step synthesis is

US 6,420,169 B1

41

```

1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0
S = 0 0 0 1 1 1 0 0 0 0 0 0 1 1 1 0 0 0
1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0
0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0
0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1

```

The round denoted by (B) places B in all products because the reaction area was uniformly activated (the mask for B consisted entirely of 1's).

The number of compounds k formed in a synthesis consisting of r rounds, in which the i th round has b_i chemical reactants and Z_i nulls, is

$$k = \sum (b_i + z_i)$$

and the number of chemical steps n is

$$n = \sum b_i$$

The number of compounds synthesized when $b=a$ (the number of chemical building blocks) and $z=0$ in all rounds is $a^{n/a}$, compared with 2^n for a binary synthesis. For $n=20$ and $a=5$, 625 compounds (all tetramers) would be formed, compared with 1.049×10^6 compounds in a binary synthesis with the same number of chemical steps.

It should also be noted that rounds in a polynomial can be nested, as in

$$(A+(B+\emptyset)(C+\emptyset))(D+\emptyset)$$

The products are AD, BCD, BD, CD, D, A, BC, B, C, and \emptyset .

Binary syntheses are attractive for two reasons. First, they generate the maximal number of products (2^n) for a given number of chemical steps (n). For four reactants, 16 compounds are formed in the binary synthesis, whereas only 4 are made when each round has two reactants. A 10-step binary synthesis yields 1,024 compounds, and a 20-step synthesis yields 1,048,576. Second, products formed in a binary synthesis are a complete nested set with lengths ranging from 0 to n . All compounds that can be formed by deleting one or more units from the longest product (the n -mer) are present. Contained within the binary set are the smaller sets that would be formed from the same reactants using any other set of masks (e.g., AC, AD, BC, and BD formed in the synthesis shown in FIG. 5 are present in the set of 16 formed by the binary synthesis). In some cases, however, the experimentally achievable spatial resolution may not suffice to accommodate all the compounds that could be formed on a single substrate. Therefore, practical limitations may require one to select a particular subset of the possible switch vectors for a given synthesis.

1. EXAMPLE

FIG. 25 illustrates a synthesis with a binary masking strategy. The binary masking strategy provides the greatest number of sequences for a given number of cycles. According to this embodiment, a mask m_1 allows illumination of half of the substrate. The substrate is then exposed to the building block A, which binds at the illuminated regions.

Thereafter, the mask m_2 allows illumination of half of the previously illuminated region, while it does not illuminate half of the previously illuminated region. The building block B is then added, which binds at the illuminated regions from m_2 .

42

The process continues with masks m_3 , m_4 , and m_5 , resulting in the product array shown in the bottom portion of the figure. The process generates 32 (2 raised to the power of the number of monomers) sequences with 5 (the number of monomers) cycles.

2. EXAMPLE

FIG. 26 illustrates another preferred binary masking strategy which is referred to herein as the gray code masking strategy. According to this embodiment, the masks m_1 to m_5 are selected such that a side of any given synthesis region is defined by the edge of only one mask. The site at which the sequence BCDE is formed, for example, has its right edge defined by m_5 and its left side formed by mask m_4 (and no other mask is aligned on the sides of this site). Accordingly, problems created by misalignment, diffusion of light under the mask and the like will be minimized.

3. EXAMPLE

FIG. 27 illustrates another binary masking strategy. According to this scheme, referred to herein as a modified gray code masking strategy, the number of masks needed is minimized. For example, the mask m_2 could be the same mask as m_1 and simply translated laterally. Similarly, the mask m_4 could be the same as mask m_3 and simply translated laterally.

4. EXAMPLE

A four-step synthesis is shown in FIG. 28A. The reactants are the ordered set {A,B,C,D}. In the first cycle, illumination through m_1 activates the upper half of the synthesis area. Building block A is then added to give the distribution 602. Illumination through mask m_2 (which activates the lower half), followed by addition of B yields the next intermediate distribution 604. C is added after illumination through m_3 (which activates the left half) giving the distribution 604, and D after illumination through m_4 (which activates the right half), to yield the final product pattern 608 {AC,AD,BC,BD}.

5. EXAMPLE

The above masking strategy for the synthesis may be extended for all 400 dipeptides from the 20 naturally occurring amino acids as shown in FIG. 28B. The synthesis consists of two rounds, with 20 photolysis and chemical coupling cycles per round. In the first cycle of round 1, mask 1 activates $1/20$ th of the substrate for coupling with the first of 20 amino acids. Nineteen subsequent illumination/coupling cycles in round 1 yield a substrate consisting of 20 rectangular stripes each bearing a distinct member of the 20 amino acids. The masks of round 2 are perpendicular to round 1 masks and therefore a single illumination/coupling cycle in round 2 yields 20 dipeptides. The 20 illumination/coupling cycles of round 2 complete the synthesis of the 400 dipeptides.

6. EXAMPLE

The power of the binary masking strategy can be appreciated by the outcome of a 10-step synthesis that produced 1,024 peptides. The polynomial expression for this 10-step binary synthesis was:

$$(f+\emptyset)(Y+\emptyset)(G+\emptyset)(A+\emptyset)(G+\emptyset)(T+\emptyset)(F+\emptyset)(L+\emptyset)(S+\emptyset)(F+\emptyset)$$

Each peptide occupied a $400 \times 400 \mu\text{m}$ square. A 32×32 peptide array (1,024 peptides, including the null peptide and

US 6,420,169 B1

43

10 peptides of $l=1$, and a limited number of duplicates) was clearly evident in a fluorescence scan following side group deprotection and treatment with the antibody 3E7 and fluoresceinated antibody. Each synthesis site was a $400 \times 400 \mu\text{m}$ square.

The scan showed a range of fluorescence intensities, from a background value of 3,300 counts to 22,400 counts in the brightest square ($x=20$, $y=9$). Only 15 compounds exhibited an intensity greater than 12,300 counts. The median value of the array was 4,800 counts.

The identity of each peptide in the array could be determined from its x and y coordinates (each range from 0 to 31) and the map of FIG. 29. The chemical units at positions 2, 5, 6, 9, and 10 are specified by the y coordinate and those at positions 1, 3, 4, 7, 8 by the x coordinate. All but one of the peptides was shorter than 10 residues. For example, the peptide at $x=12$ and $y=3$ is YGAGF (SEQ ID NO:30; positions 1, 6, 8, 9, and 10 are nulls). YGAFLS (SEQ ID NO:4), the brightest element of the array, is at $x=20$ and $y=9$.

It is often desirable to deduce a binding affinity of a given peptide from the measured fluorescence intensity. Conceptually, the simplest case is one in which a single peptide binds to a univalent antibody molecule. The fluorescence scan is carried out after the slide is washed with buffer for a defined time. The order of fluorescence intensities is then a measure primarily of the relative dissociation rates of the antibody-peptide complexes. If the on-rate constants are the same (e.g., if they are diffusion-controlled), the order of fluorescence intensities will typically correspond to the order of binding affinities. However, the situation is sometimes more complex because a bivalent primary antibody and a bivalent secondary antibody are used. The density of peptides in a synthesis area corresponded to a mean separation of ~ 7 nm, which would allow multivalent antibody-peptide interactions. Hence, fluorescence intensities obtained according to the method herein will often be a qualitative indicator of binding affinity. For a more complete analysis of how the present invention can be extended to the binding affinity of an immobilized ligand to a receptor, see U.S. Ser. No. 07/796,947, filed Nov. 22, 1991, now U.S. Pat. No. 5,324,633 and incorporated herein by reference.

Another important consideration is the fidelity of synthesis. Deletions are produced by incomplete photodeprotection or incomplete coupling. The coupling yield per cycle in these experiments is typically between 85% and 95%. The contribution to the net coupling yield from photodeprotection and chemical coupling has been assessed in the following ways. The photolysis rate for NVOC-amino acids was experimentally determined and illumination conditions that ensure greater than 99% of the amino acids have been photodeprotected were chosen. The chemical coupling efficiency of selected amino acids on substrates employed in this invention has also been determined. For example, in order to determine the coupling efficiency of Leu to Leu, NVOC was first selectively photolyzed from one region of a NVOC-Leu derivitized surface. The photochemically deprotected amino groups in this region were then coupled to a FMOC-Leu-OBt. At this stage, incomplete Leu to Leu coupling would leave unreacted amino groups. A second photolysis step was then used to photolyze a different region of the substrate. Treatment of the substrate with FITC would label the free amino groups that remain from incomplete chemical coupling in the first region and free amino groups exposed by photolysis in the second region. Direct comparison of the quantitative fluorescence signal from both regions indicates the extent of chemical coupling. If the chemical coupling yield is high, the ratio of the signals of the first to

44

the second photolysis regions is low. This technique has been used in order to develop the experimental conditions that maximize chemical coupling.

Implementing the switch matrix by masking is imperfect because of light diffraction, internal reflection, and scattering. Consequently, stowaways (chemical units that should not be on board) arise by unintended illumination of regions that should be dark. A binary synthesis array contains many of the controls needed to assess the fidelity of a synthesis. For example, the fluorescence signal from a synthesis area nominally containing a tetrapeptide ABCD could come from a tripeptide deletion impurity such as ACD. Such an artifact would be ruled out by the finding that the fluorescence intensity of the ACD site is less than that of the ABCD site.

The fifteen most highly fluorescent peptides in the array obtained with the synthesis of 1,024 peptides described above, were YGAFLS (SEQ ID NO:4), YGAFLS (SEQ ID NO:5), YGAFL (SEQ ID NO:6), YGGFLS (SEQ ID NO:7), YGAF (SEQ ID NO:8), YGALS (SEQ ID NO:9), YGGFS (SEQ ID NO:10), YGAL (SEQ ID NO:11), YGAFLF (SEQ ID NO:12), YGAF (SEQ ID NO:8), YGAFF (SEQ ID NO:13), YGGLS (SEQ ID NO:14), YGGFL (SEQ ID NO:15) and SEQ ID NO:15), YGAFFS (SEQ ID NO:16), and YGAFLSF (SEQ ID NO:17). A striking feature is that all fifteen begin with YG, which agrees with previous work showing that an amino-terminal tyrosine is a key determinant of binding to antibody 3E7. Residue 3 of this set is either A or G, and residue 4 is either F or L. The exclusion of S and T from these positions is clear cut. The finding that the preferred sequence is YG (A/G) (F/L) fits nicely with the outcome of a study in which a very large library of peptides on phage generated by recombinant DNA methods was screened for binding to antibody 3E7 (see Cwirla et al., *Proc. Natl. Acad. Sci. USA*, (1990) 87:6378, incorporated herein by reference). Additional binary syntheses based on leads from peptides on phage experiments show that YGAFMQ (SEQ ID NO:18), YGAFM (SEQ ID NO:19), and YGAFQ (SEQ ID NO:20) give stronger fluorescence signals than does YGGFM (SEQ ID NO:21), the immunogen used to obtain antibody 3E7.

Variations on the above masking strategy will be valuable in certain circumstances. For example, if a "kernel" sequence of interest consists of PQR separated from XYZ, the aim is to synthesize peptides in which these units are separated by a variable number of different residues. The kernel can be placed in each peptide by using a mask that has 1's everywhere. The polynomial representation of a suitable synthesis is:

$$(P)(Q)(R)(A+\emptyset)(B+\emptyset)(C+\emptyset)(D+\emptyset)(X)(Y)(Z)$$

Sixteen peptides will be formed, ranging in length from the 6-mer PQRXYZ to the 10-mer PQRABCDXYZ.

Several other masking strategies will also find value in selected circumstances. By using a particular mask more than once, two or more reactants will appear in the same set of products. For example, suppose that the mask for an 8-step synthesis is

A 11110000
B 00001111
C 11001100
D 00110011
E 10101010
F 01010101
G 11110000
H 00001111

The products are ACEG, ACFG, ADEG, ADFG, BCEH, BCFH, BDEH, and BDFH. A and G always appear in the

same product, although not necessarily next to each other, because their additions were directed by the same mask, and likewise for B and H.

7. EXAMPLE

The synthesis strategies shown above are useful in many different applications. To aid in applying the present invention to any desired synthesis, the following illustrative example is provided. Assume one wishes to synthesize polymers up to 4 monomers in length. A given polymer can be designated as $Y_1 Y_2 Y_3 Y_4$. If the monomer set contains 20 members, then the set (5) can be represented as follows:

$$S=\{M_1, M_2, M_3, \dots M_{20}\},$$

where, for example, Y_1 may be M_1 , Y_2 may be M_{16} , etc. Then, if one desires to synthesize all polymers in which each position is varied through the entire set of monomers, then the synthesis can be represented as:

$$\text{Product of Synthesis}=\text{SY}_2\text{Y}_3\text{Y}_4=\text{Y}_1\text{SY}_3\text{Y}_4+\text{Y}_1\text{Y}_2\text{SY}_4+\text{Y}_1\text{Y}_2\text{Y}_3\text{S}.$$

In the above polynomial, there are four terms, and each term represents 20 different compounds. If one desires to synthesize all polymers in which two positions are varied through the entire set of monomers, then the synthesis can be represented as:

$$\text{Product of Synthesis}=\text{SSY}_3\text{Y}_4+\text{SY}_2\text{SY}_4+\text{SY}_2\text{Y}_3\text{S}+\text{Y}_1\text{SSY}_4+\text{Y}_1\text{SY}_3\text{S}+\text{Y}_1\text{Y}_2\text{SS}.$$

In the above polynomial, there are six terms, and each term represents 400 compounds. If one desires to synthesize all polymers in which three positions are varied through the entire set of monomers, then the synthesis can be represented as:

$$\text{Product of Synthesis}=\text{SSSY}_4+\text{SSY}_3\text{S}=\text{SY}_2\text{SS}+\text{Y}_1\text{SSS}.$$

In the above polynomial, there are four terms and each term represents 8,000 compounds. If one desires to synthesize all polymers in which four positions are varied through the entire set of monomers, then the synthesis can be represented as:

$$\text{Product of Synthesis}=\text{SSSS}$$

In the above polynomial, there is one term, which represents 160,000 compounds.

By modeling the synthesis as a polynomial expression, one can more easily discern the appropriate masking strategy required to effect the synthesis.

8. EXAMPLE

One example of the power of this strategy involved the mapping of a binding epitope on dynorphin B. The sequence of this epitope was demonstrated (as shown below) to be RQFKVVT (SEQ. ID NO:31). An array of peptides synthesized using the general protection-deprotection technology outlined above. Referring to FIG. 30, the following peptides were synthesized: row 1=RXFKVVT; row 2=RQXFKVVT; row 3=RQXKVVT; and row 4=RXKVVT. In each row, "X" represents a group of four amino acids, which were simultaneously added to the immobilized peptide. The particular group of amino acids used as "X" on a given block are identified by column number as follows: column 1 [G,A,R,K in a ratio 1:1:1:1]; column 2 [null]; column 3 [C,M,S,T in the ratio 1:1:1:5]; column 4 [null]; column 5 [F,Y,W,H in the ratio 1:1:1:1]; column 6 [null]; column 7 [D,E,N,Q in the

ratio 1:1:1:1]; column 8 [null]; column 9 [V,L,I,P in the ratio 2:1:4:1]; and column 10 [null].

After these peptides were synthesized, they were screened with an anti-dynorphin B murine monoclonal antibody and then exposed to a fluorescently labeled goat anti-mouse antibody. The brightest region (column 2, row 2) corresponds to the binding epitope, RQFKVVT. Other strongly labeled regions, include column 7, row 1 and column 5, row 3, each of which contains some peptides having the above binding epitope. It should be noted that the varying ratios of some of the amino acids within a given group were necessary to obtain roughly equal binding efficiencies among the four amino acids.

9. EXAMPLE

A much larger array of compounds was prepared as shown in FIG. 31. In the first four steps, KVVT (SEQ. ID NO:36) was synthesized over the entire substrate. Next, F was added over 50% of the substrate via a column mask as displayed in step 5. The next 20 steps involved addition of a row of each of the genetically-coded amino acids in a stepwise fashion down the substrate. One-twentieth of the substrate was exposed on each pass. Next, Q was added over one-half of the entire substrate as shown at step 26. Finally, R was added over the entire substrate.

Four classes of peptide were produced: (1) RYKVVT (SEQ. ID NO:32), (2) RQYKVVT (SEQ. ID NO:33), (3) RQYFKVVT (SEQ. ID NO:34), and (4) RYFKVVT (SEQ. ID NO:35). In each case Y represents all 20 L-amino acids. FIG. 32 is an image of a fluorescence scan prepared after the final array of peptides was exposed to anti-dynorphin B mouse monoclonal antibody followed by goat anti-mouse antibody. The top one-fourth of the image was prepared according to the synthesis procedure outlined above.

10. EXAMPLE

The binding epitope on dynorphin B for the anti-dynorphin B antibody described above was determined by deleting various amino acids and combinations of amino acids from the overall sequence of dynorphin B. The peptides containing these various deletions were prepared on a substrate by the general protection-deprotection methods described above. Then the substrate was exposed to the anti-dynorphin B monoclonal antibody and a fluorescence image was produced as shown in FIG. 33. Each quadrant of the image contained at least one site for each of the subject peptides (identified in Table 7).

From the intensity distribution contained in this plot, it was possible to determine the relative binding affinity to various of the peptides. This information is summarized in Table 7 below where N is the number of sites of which the peptide was synthesized.

TABLE 7

SEQUENCE ADJUSTED	N	RELATIVE INTENSITY
F	8	4.9 ± 3.2
F L R	8	4.7 ± 2.1
F L R R	4	6.0 ± 2.7
F L R R Q	4	8.2 ± 2.6
F L R R Q F	4	8.5 ± 3.4
F L R R Q F K	4	10.2 ± 2.8
F L R R Q F K V	8	13.7 ± 3.0
F L R R Q F K V V	4	37.4 ± 14.6 (30.5 ± 5.4)
F L R R Q F K V V T	4	84.2 ± 28.2 (98.3 ± 1.2)

TABLE 7-continued

SEQUENCE ADJUSTED	N	RELATIVE INTENSITY
L R R Q F K V V T	4	86.9 ± 13.7
R R Q F K V V T	4	98.8 ± 0.9
R Q F K V V T	8	93.6 ± 10.0 (96.2 ± 7.1)
Q F K V V T	4	36.2 ± 16.0
F K V V T	8	12.9 ± 3.7
K V V T	4	10.7 ± 1.8
V V T	4	8.2 ± 1.2
V T	8	7.9 ± 3.0
T	4	7.0 ± 3.9

FIG. 34 is a bar graph showing the relative binding affinities for each of the peptides.

C. Linker Selection

According to preferred embodiments the linker molecules used as an intermediary between the synthesized polymers and the substrate are selected for optimum length and/or type for improved binding interaction with a receptor. According to this aspect of the invention diverse linkers of varying length and/or type are synthesized for subsequent attachment of a ligand. Through variations in the length and type of linker, it becomes possible to optimize the binding interaction between an immobilized ligand and its receptor.

The degree of binding between a ligand (peptide, inhibitor, hapten, drug, etc.) and its receptor (enzyme, antibody, etc.) when one of the partners is immobilized onto a substrate will in some embodiments depend on the accessibility of the receptor in solution to the immobilized ligand. The accessibility in turn will depend on the length and/or type of linker molecule employed to immobilize one of the partners. Preferred embodiments of the invention therefore employ the VLSIPS™ synthesis technique described herein to generate an array of, preferably, inactive or inert linkers of varying length and/or type, using photochemical protecting groups to selectively expose different regions of the substrate and to build upon chemically-active groups.

In the simplest embodiment of this concept, the same unit is attached to the substrate in varying multiples or lengths in known locations on the substrate via VLSIPS™ synthesis techniques to generate an array of polymers of varying length. A single ligand (peptide, drug, hapten, etc.) is attached to each of them, and an assay is performed with the binding site to evaluate the degree of binding with a receptor that is known to bind to the ligand. In cases where the linker length impacts the ability of the receptor to bind to the ligand, varying levels of binding will be observed. In general, the linker which provides the highest binding will then be used to assay other ligands synthesized in accordance with the techniques herein.

According to other embodiments the binding between a single ligand/receptor pair is evaluated for linkers of diverse monomer sequence. According to these embodiments, the linkers are synthesized in an array in accordance with the techniques herein and have different monomer sequences (and, optionally, different lengths). Thereafter, all of the linker molecules are provided with a ligand known to have at least some binding affinity for a given receptor. The given receptor is then exposed to the ligand and binding affinity is deduced. Linker molecules which provide adequate binding between the ligand and receptor are then utilized in screening studies.

D. Protecting Groups

As discussed above, selectively removable protecting groups allow creation of well defined areas of substrate

surface having differing reactivities. Preferably, the protecting groups are selectively removed from the surface by applying a specific activator, such as electromagnetic radiation of a specific wavelength and intensity. More preferably, the specific activator exposes selected areas of the surface to remove the protecting groups in the exposed areas.

Protecting groups of the present invention are used in conjunction with solid phase oligomer syntheses, such as peptide syntheses using natural or unnatural amino acids, nucleotide syntheses using deoxyribonucleic and ribonucleic acids, oligosaccharide syntheses, and the like. In addition to protecting the substrate surface from unwanted reaction, the protecting groups block a reactive end of the monomer to prevent self-polymerization. For instance, attachment of a protecting group to the amino terminus of an activated amino acid, such as an N-hydroxysuccinimide-activated ester of the amino acid, prevents the amino terminus of one monomer from reacting with the activated ester portion of another during peptide synthesis. Alternatively, the protecting group may be attached to the carboxyl group of an amino acid to prevent reaction at this site. Most protecting groups can be attached to either the amino or the carboxyl group of an amino acid, and the nature of the chemical synthesis will dictate which reactive group will require a protecting group. Analogously, attachment of a protecting group to the 5'-hydroxyl group of a nucleoside during synthesis using for example, phosphate-triester coupling chemistry, prevents the 5'-hydroxyl of one nucleoside from reacting with the 3'-activated phosphate-triester of another.

Regardless of the specific use, protecting groups are employed to protect a moiety on a molecule from reacting with another reagent. Protecting groups of the present invention have the following characteristics: they prevent selected reagents from modifying the group to which they are attached; they are stable (that is, they remain attached to the molecule) to the synthesis reaction conditions; they are removable under conditions that do not adversely affect the remaining structure; and once removed, they do not react appreciably with the surface or surface-bound oligomer. The selection of a suitable protecting group will depend, of course, on the chemical nature of the monomer unit and oligomer, as well as the specific reagents they are to protect against.

In a preferred embodiment, the protecting groups are photoactivatable. The properties and uses of photoreactive protecting compounds have been reviewed. See, McCray et al., *Ann. Rev. of Biophys. and Biophys. Chem.* (1989) 18:239-270, which is incorporated herein by reference. Preferably, the photosensitive protecting groups will be removable by radiation in the ultraviolet (UV) or visible portion of the electromagnetic spectrum. More preferably, the protecting groups will be removable by radiation in the near UV or visible portion of the spectrum. In some embodiments, however, activation may be performed by other methods such as localized heating, electron beam lithography, laser pumping, oxidation or reduction with microelectrodes, and the like. Sulfonyl compounds are suitable reactive groups for electron beam lithography. Oxidative or reductive removal is accomplished by exposure of the protecting group to an electric current source, preferably using microelectrodes directed to the predefined regions of the surface which are desired for activation. Other methods may be used in light of this disclosure.

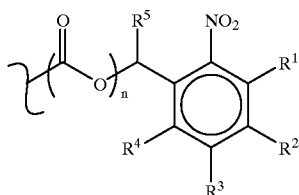
Many, although not all, of the photoremovable protecting groups will be aromatic compounds that absorb near-UV and visible radiation. Suitable photoremovable protecting

US 6,420,169 B1

49

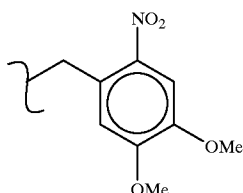
groups are described in, for example, McCray et al., Patchornik, *J. Amer. Chem. Soc.* (1970) 92:6333, and Amit et al., *J. Org. Chem.* (1974) 39:192, which are incorporated herein by reference.

A preferred class of photoremovable protecting groups has the general formula:

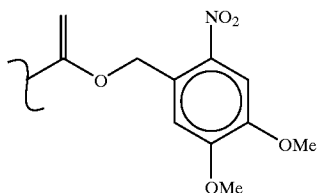


where R^1 , R^2 , R^3 , and R^4 independently are a hydrogen atom, a lower alkyl, aryl, benzyl, halogen, hydroxyl, alkoxyl, thiol, thioether, amino, nitro, carboxyl, formate, formamido or phosphido group, or adjacent substituents (i.e., R^1 - R^2 , R^2 - R^3 , R^3 - R^4) are substituted oxygen groups that together form a cyclic acetal or ketal; R^5 is a hydrogen atom, an alkoxyl, alkyl, halo, aryl, or alkenyl group, and $n=0$ or 1.

A preferred protecting group, 6-nitroveratryl (NV), which is used for protecting the carboxyl terminus of an amino acid or the hydroxyl group of a nucleotide, sugar, or carbohydrate for example, is formed when R^2 and R^3 are each a methoxy group, R^1 , R^4 and R^5 are each a hydrogen atom, and $n=0$:

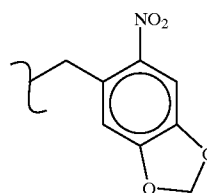


A preferred protecting group, 6-nitroveratryloxycarbonyl (NVOC), which is used to protect the amino terminus of an amino acid, or the hydroxyl group of a nucleotide, sugar, or carbohydrate for example, is formed when R^2 and R^3 are each a methoxy group, R^1 , R^4 and R^5 are each a hydrogen atom, and $n=1$:

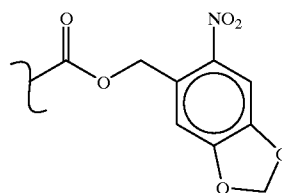


Another preferred protecting group, 6-nitropiperonyl (NP), which is used for protecting the carboxyl terminus of an amino acid or the hydroxyl group of a nucleotide, sugar, or carbohydrate for example, is formed when R^2 and R^3 together form a methylene acetal, R^1 , R^4 and R^5 are each a hydrogen atom, and $n=0$:

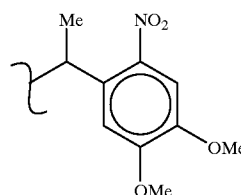
50



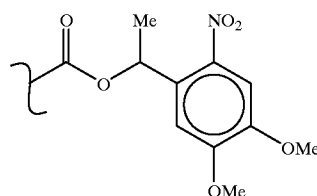
Another preferred protecting group, 6-nitropiperonyloxycarbonyl (NPOC), which is used to protect the amino terminus of an amino acid, or the hydroxyl group of a nucleotide, sugar, or carbohydrate for example, is formed when R^2 and R^3 together form a methylene acetal, R^1 , R^4 and R^5 are each a hydrogen atom, and $n=1$:



A most preferred protecting group, methyl-6-nitroveratryl (MeNV), which is used for protecting the carboxyl terminus of an amino acid or the hydroxyl group of a nucleotide, sugar or carbohydrate for example, is formed when R^2 and R^3 are each a methoxy group, R^1 and R^4 are each a hydrogen atom, R^5 is a methyl group, and $n=0$:



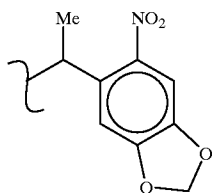
Another most preferred protecting group, methyl-6-nitroveratryloxycarbonyl (MeNVOC), which is used to protect the amino terminus of an amino acid, or the hydroxyl group of a nucleotide, sugar, or carbohydrate for example, is formed when R^2 and R^3 are each a methoxy group, R^1 and R^4 are each a hydrogen atom, R^5 is a methyl group, and $n=1$:



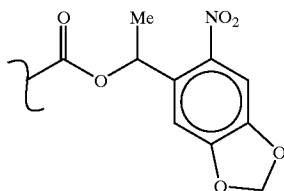
Another most preferred protecting group, methyl-6-nitropiperonyl (MeNP), which is used for protecting the carboxyl terminus of an amino acid or the hydroxyl group of a nucleotide, sugar or carbohydrate for example, is formed when R^2 and R^3 together form a methylene acetal, R^1 and R^4 are each a hydrogen atom, R^5 is a methyl group, and $n=0$:

US 6,420,169 B1

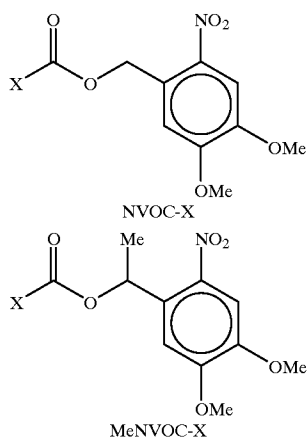
51



Another most preferred protecting group, methyl-6-nitropiperonyloxycarbonyl (MeNPOC), which is used to protect the amino terminus of an amino acid or to protect the 5' hydroxyl of nucleosides, nucleotides, carbohydrates, or sugars for example, is formed when R² and R³ together form a methylene acetal, R¹ and R⁴ are each a hydrogen atom, R⁵ is a methyl group, and n=1:



A protected amino acid having a photoactivatable oxycarbonyl protecting group, such as NVOC or NPOC or their corresponding methyl derivatives, MeNVOC or MeNPOC, respectively, on the amino terminus is formed by acylating the amine of the amino acid or 5' hydroxyl of a nucleotide, sugar or carbohydrate with an activated oxycarbonyl ester of the protecting group. Examples of activated oxycarbonyl esters of NVOC and MeNVOC have the general formula:

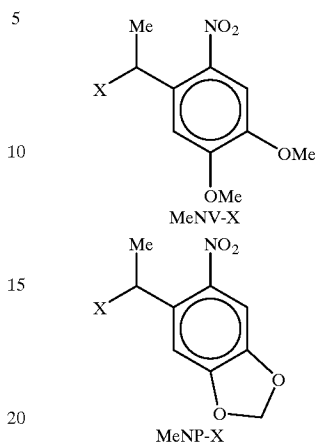


where X is halogen, mixed anhydride, phenoxy, p-nitrophenoxy, N-hydroxysuccinimide, and the like.

A protected amino acid or nucleotide having a photoactivatable protecting group, such as NV or NP or their corresponding methyl derivatives, MeNV or MeNP, respectively, on the carboxy terminus of the amino acid or 5'-hydroxy terminus of the nucleotide, is formed by acylating the carboxy terminus or 5'-OH with an activated benzyl

52

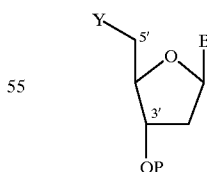
derivative of the protecting group. Examples of activated benzyl derivatives of MeNV and MeNP have the general formula:



where X is halogen, hydroxyl, tosyl, mesyl, trifluoromethyl, diazo, azido, and the like.

Another method for generating protected monomers is to react the benzylic alcohol derivative of the protecting group with an activated ester of the monomer. For example, to protect the carboxyl terminus of an amino acid, an activated ester of the amino acid is reacted with the alcohol derivative of the protecting group, such as 6-nitroveratrol (NVOH). Examples of activated esters suitable for such uses include halo-formate, mixed anhydride, imidazolyl formate, acyl halide, and also include formation of the activated ester in situ the use of common reagents such as DCC and the like. See Atherton et al. for other examples of activated esters.

A further method for generating protected monomers is to react the benzylic alcohol derivative of the protecting group with an activated carbon of the monomer. For example, to protect the 5'-hydroxyl group of a nucleic acid, a derivative having a 5'-activated carbon is reacted with the alcohol derivative of the protecting group, such as methyl-6-nitropiperonol (MeNPOH). Examples of nucleotides having activating groups attached to the 5'-hydroxyl group have the general formula:

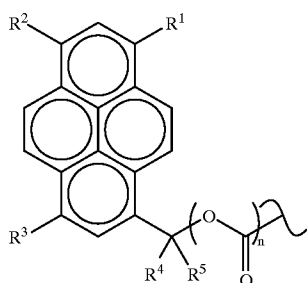


where Y is a halogen atom, a tosyl, mesyl, trifluoromethyl, azido, or diazo group, and the like.

Another class of preferred photochemical protecting groups has the formula:

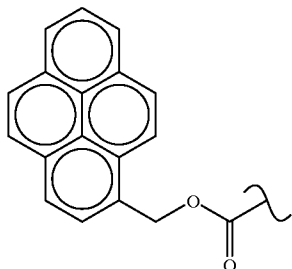
US 6,420,169 B1

53

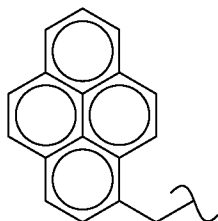


where R^1 , R^2 , and R^3 independently are a hydrogen atom, a lower alkyl, aryl, benzyl, halogen, hydroxyl, alkoxy, thiol, thioether, amino, nitro, carboxyl, formate, formamido, sulfanates, sulfido or phosphido group, R^4 and R^5 independently are a hydrogen atom, an alkoxy, alkyl, halo, aryl, or alkenyl group, and $n=0$ or 1.

A preferred protecting group, 1-pyrenylmethyloxycarbonyl (PyROC), which is used to protect the amino terminus of an amino acid, or the hydroxyl group of nucleotide, sugar or carbohydrate for example, is formed when R^1 through R^5 are each a hydrogen atom and $n=1$:

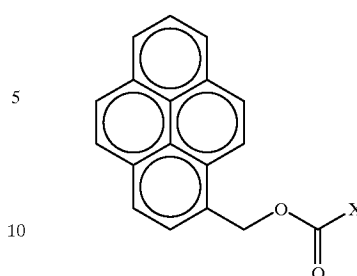


Another preferred protecting group, 1-pyrenylmethyl (PyR), which is used for protecting the carboxy terminus of an amino acid or the hydroxyl group of a nucleotide, sugar or carbohydrate for example, is formed when R^1 through R^5 are each a hydrogen atom and $n=0$:



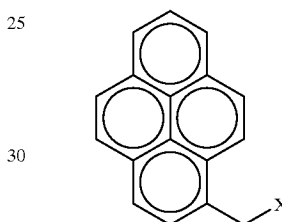
An amino acid having a pyrenylmethyloxycarbonyl protecting group on its amino terminus is formed by acylation of the free amine of amino acid with an activated oxycarbonyl ester of the pyrenyl protecting group. Examples of activated oxycarbonyl esters of PyROC have the general formula:

54



where X is halogen, or mixed anhydride, p-nitrophenoxy, or N-hydroxysuccinimide group, and the like.

A protected amino acid or nucleotide having a photoactivatable protecting group, such as PyR, on the carboxy terminus of the amino acid or 5'-hydroxy terminus of the nucleic acid, respectively, is formed by acylating the carboxy terminus or 5'-OH with an activated pyrenylmethyl derivative of the protecting group. Examples of activated pyrenylmethyl derivatives of PyROC have the general formula:



where X is a halogen atom, a hydroxyl, diazo, or azido group, and the like.

Another method of generating protected monomers is to react the pyrenylmethyl alcohol moiety of the protecting group with an activated ester of the monomer. For example, an activated ester of an amino acid can be reacted with the alcohol derivative of the protecting group, such as pyrenylmethyl alcohol (PyROH), to form the protected derivative of the carboxy terminus of the amino acid. Examples of activated esters include halo-formate, mixed anhydride, imidazolyl formate, acyl halide, and also include formation of the activated ester in situ and the use of common reagents such as DCC and the like.

Clearly, many photosensitive protecting groups are suitable for use in the present invention.

In preferred embodiments, the substrate is irradiated to remove the photoremovable protecting groups and create regions having free reactive moieties and side products resulting from the protecting group. The removal rate of the protecting groups depends on the wavelength and intensity of the incident radiation, as well as the physical and chemical properties of the protecting group itself. Preferred protecting groups are removed at a faster rate and with a lower intensity of radiation. For example, at a given set of conditions, MeNVOC and MeNPOC are photolytically removed from the N-terminus of a peptide chain faster than their unsubstituted parent compounds, NVOC and NPOC, respectively.

Removal of the protecting group is accomplished by irradiation to separate the reactive group and the degradation products derived from the protecting group. Not wishing to be bound by theory, it is believed that irradiation of an NVOC- and MeNVOC-protected oligomers occurs by the following reaction schemes:

US 6,420,169 B1

55

NVOC-AA → 3,4-dimethoxy-6-nitrosobenzaldehyde + CO₂ + AAMeNVOC-AA → 3,4-dimethoxy-6-nitrosoacetophenone + CO₂ + AA

where AA represents the N-terminus of the amino acid oligomer.

Along with the unprotected amino acid, other products are liberated into solution: carbon dioxide and a 2,3-dimethoxy-6-nitrosophenylcarbonyl compound, which can react with nucleophilic portions of the oligomer to form unwanted secondary reactions. In the case of an NVOC-protected amino acid, the degradation product is a nitrosobenzaldehyde, while the degradation product for the other is a nitrosophenyl ketone. For instance, it is believed that the product aldehyde from NVOC degradation reacts with free amines to form a Schiff base (imine) that affects the remaining polymer synthesis. Preferred photoremovable protecting groups react slowly or reversibly with the oligomer on the support.

Again not wishing to be bound by theory, it is believed that the product ketone from irradiation of a MeNVOC-protected oligomer reacts at a slower rate with nucleophiles on the oligomer than the product aldehydes from irradiation of the same NVOC-protected oligomer. Although not unambiguously determined, it is believed that this difference in reaction rate is due to the difference in general reactivity between aldehydes and ketones towards nucleophiles due to steric and electronic effects.

The photoremovable protecting groups of the present invention are readily removed. For example, the photolysis of N-protected L-phenylalanine in solution having different photoremovable protecting groups was analyzed, and the results are presented in the following table:

TABLE 9

Photolysis of Protected L-Phe-OH				
Solvent	t _{1/2} in seconds			
	NBOC	NVOC	MeNVOC	MeNPOC
Dioxane	1288	110	24	19
5 mM H ₂ SO ₄ /Dioxane	1575	98	33	22

The half life, t_{1/2}, is the time in seconds required to remove 50% of the starting amount of protecting group. NBOC is the 6-nitrobenzyloxycarbonyl group, NVOC is the 6-nitroveratryloxycarbonyl group, MeNVOC is the methyl-6-nitroveratryloxycarbonyl group, and MeNPOC is the methyl-6-nitropiperonyloxycarbonyl group. The concentration of each protected phenylalanine was 0.10 mM.

Table 9 shows that deprotection of NVOC-, MeNVOC-, and MeNPOC-protected phenylalanine proceeded faster than the deprotection of NBOC. Furthermore, it shows that the deprotection of the two derivatives that are substituted on the benzylic carbon, MeNVOC and MeNPOC, were photolyzed at the highest rates in both dioxane and acidified dioxane.

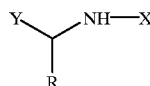
1. Use of Photoremovable Groups During Solid-Phase Synthesis of Peptides

The formation of peptides on a solid-phase support requires the stepwise attachment of an amino acid to a substrate-bound growing chain. In order to prevent unwanted polymerization of the monomeric amino acid under the reaction conditions, protection of the amino terminus of the amino acid is required. After the monomer is coupled to the end of the peptide, the N-terminal protecting group is removed, and another amino acid is coupled to the

56

chain. This cycle of coupling and deprotecting is continued for each amino acid in the peptide sequence. See Merrifield, *J. Am. Chem. Soc.* (1963) 85:2149, and Atherton et al., "Solid Phase Peptide Synthesis" 1989, IRL Press, London, both incorporated herein by reference for all purposes. As described above, the use of a photoremovable protecting group allows removal of selected portions of the substrate surface, via patterned irradiation, during the deprotection cycle of the solid phase synthesis. This selectively allows spatial control of the synthesis—the next amino acid is coupled only to the irradiated areas.

In one embodiment, the photoremovable protecting groups of the present invention are attached to an activated ester of an amino acid at the amino terminus:



where R is the side chain of a natural or unnatural amino acid, X is a photoremovable protecting group, and Y is an activated carboxylic acid derivative. The photoremovable protecting group, X, is preferably NVOC, NPOC, PyROC, MeNVOC, MeNPOC, and the like as discussed above. The activated ester, Y, is preferably a reactive derivative having a high coupling efficiency, such as an acyl halide, mixed anhydride, N-hydroxysuccinimide ester, HOBt ester perfluorophenyl ester, or urethane protected acid, and the like. Other activated esters and reaction conditions are well known (See Atherton et al.).

2. Use of Photoremovable Protecting Groups During Solid-Phase Synthesis of Oligonucleotides

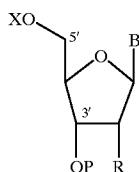
The formation of oligonucleotides on a solid-phase support requires the stepwise attachment of a nucleotide to a substrate-bound growing oligomer. In order to prevent unwanted polymerization of the monomeric nucleotide under the reaction conditions, protection of the 5'-hydroxyl group of the nucleotide is required. After the monomer is coupled to the end of the oligomer, the 5'-hydroxyl protecting group is removed, and another nucleotide is coupled to the chain. This cycle of coupling and deprotecting is continued for each nucleotide in the oligomer sequence. See Gait, "Oligonucleotide Synthesis: A Practical Approach" 1984, IRL Press, London, incorporated herein by reference for all purposes. As described above, the use of a photoremovable protecting group allows removal, via patterned irradiation, of selected portions of the substrate surface during the deprotection cycle of the solid phase synthesis. This selectively allows spatial control of the synthesis—the next nucleotide is coupled only to the irradiated areas.

Oligonucleotide synthesis generally involves coupling an activated phosphorous derivative on the 3'-hydroxyl group of a nucleotide with the 5'-hydroxyl group of an oligomer bound to a solid support. Two major chemical methods exist to perform this coupling: the phosphate-triester and phosphoramidite methods (See Gait). Protecting groups of the present invention are suitable for use in either method.

In a preferred embodiment, a photoremovable protecting group is attached to an activated nucleotide on the 5'-hydroxyl group:

US 6,420,169 B1

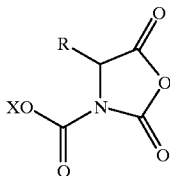
57



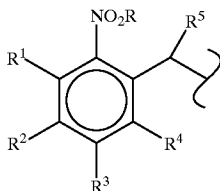
where B is the base attached to the sugar ring; R is a hydrogen atom when the sugar is deoxyribose or R is a hydroxyl group when the sugar is ribose; P represents an activated phosphorous group; and X is a photoremovable protecting group. The photoremovable protecting group, X, is preferably NV, NP, PyR, MeNV, MeNP, NVOC, NPOC, PyROC, MeNVOC, MeNPOC, and the like as described above. The activated phosphorous group, P, is preferably a reactive derivative having a high coupling efficiency, such as a phosphate-triester, phosphoramidite or the like. Other activated phosphorous derivatives, as well as reaction conditions, are well known (See Gait).

E. Amino Acid N-Carboxy Anhydrides Protected With a Photoremovable Group

During Merrifield peptide synthesis, an activated ester of one amino acid is coupled with the free amino terminus of a substrate-bound oligomer. Activated esters of amino acids suitable for the solid phase synthesis include halo-formate, mixed anhydride, imidazolyl formate, acyl halide, and also includes formation of the activated ester in situ and the use of common reagents such as DCC and the like (See Atherton et al.). A preferred protected and activated amino acid has the general formula:



where R is the side chain of the amino acid and X is a photoremovable protecting group. This compound is a urethane-protected amino acid having a photoremovable protecting group attached to the amine. A more preferred activated amino acid is formed when the photoremovable protecting group has the general formula:



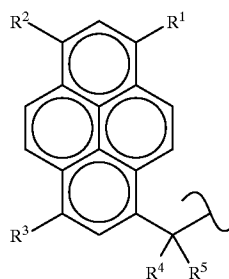
where R^1 , R^2 , R^3 , and R^4 independently are a hydrogen atom, a lower alkyl, aryl, benzyl, halogen, hydroxyl, alkoxy, thiol, thioether, amino, nitro, carboxyl, formate, formamido or phosphido group, or adjacent substituents (i.e., R^1-R^2 , R^2-R^3 , R^3-R^4) are substituted oxygen groups that together form a cyclic acetal or ketal; and R^5 is a hydrogen atom, alkoxy, alkyl, halo, aryl, or alkenyl group.

A preferred activated amino acid is formed when the photoremovable protecting group is 6-nitroveratryloxycarbonyl. That is, R^1 and R^4 are each a

58

hydrogen atom, R^2 and R^3 are each a methoxy group, and R^5 is a hydrogen atom. Another preferred activated amino acid is formed when the photoremovable group is 6-nitropiperonyl: R^1 and R^4 are each a hydrogen atom, R^2 and R^3 together form a methylene acetal, and R^5 is a hydrogen atom. Other protecting groups are possible. Another preferred activated ester is formed when the photoremovable group is methyl-6-nitroveratryl or methyl-6-nitropiperonyl.

Another preferred activated amino acid is formed when the photoremovable protecting group has the general formula:



where R^1 , R^2 , and R^3 independently are a hydrogen atom, a lower alkyl, aryl, benzyl, halogen, hydroxyl, alkoxy, thiol, thioether, amino, nitro, carboxyl, formate, formamido, sulfanate, sulfido or phosphido group, and R^4 and R^5 independently are a hydrogen atom, an alkoxy, alkyl, halo, aryl, or alkenyl group. The resulting compound is a urethane-protected amino acid having a pyrenylmethyloxycarbonyl protecting group attached to the amine. A more preferred embodiment is formed when R^1 through R^5 are each a hydrogen atom.

The urethane-protected amino acids having a photoremovable protecting group of the present invention are prepared by condensation of an N-protected amino acid with an acylating agent such as an acyl halide, anhydride, chloroformate and the like (See Fuller et al., U.S. Pat. No. 4,946,942 and Fuller et al., *J. Amer. Chem. Soc.* (1990) 112:7414-7416, both herein incorporated by reference for all purposes).

Urethane-protected amino acids having photoremovable protecting groups are generally useful as reagents during solid-phase peptide synthesis, and because of the spatial selectivity possible with the photoremovable protecting groups, are especially useful for the spatially addressable peptide synthesis. These amino acids are difunctional: the urethane group first serves to activate the carboxy terminus for reaction with the amine bound to the surface, and, once the peptide bond is formed, the photoremovable protecting group protects the newly formed amino terminus from further reaction. These amino acids are also highly reactive to nucleophiles, such as deprotected amines on the surface of the solid support, and due to this high reactivity, the solid-phase peptide coupling times are significantly reduced, and yields are typically higher.

IV. Data Collection

A. Data Collection System

Substrates prepared in accordance with the above description are used in one embodiment to determine which of the plurality of sequences thereon bind to a receptor of interest. FIG. 10 illustrates one embodiment of a device used to detect regions of a substrate which contain fluorescent markers. This device would be used, for example, to detect

US 6,420,169 B1

59

the presence or absence of a fluorescently labeled receptor such as an antibody which has bound to a synthesized polymer on a substrate.

Light is directed at the substrate from a light source **1002** such as a laser light source of the type well known to those of skill in the art such as a model no. 2025 made by Spectra Physics. Light from the source is directed at a lens **1004** which is preferably a cylindrical lens of the type well known to those of skill in the art. The resulting output from the lens **1004** is a linear beam rather than a spot of light. Thus, data can be detected substantially simultaneously along a linear array of pixels rather than on a pixel-by-pixel basis. It will be understood that while a cylindrical lens is used herein as an illustration of one technique for generating a linear beam of light on a surface, other techniques could also be utilized.

The beam from the cylindrical lens is passed through a dichroic mirror or prism and directed at the surface of the suitably prepared substrate **1008**. Substrate **1008** is placed on an x-y translation stage **1009** such as a model no. PM500-8 made by Newport. Certain locations on the substrate will fluoresce and fluorescence will be transmitted along the path indicated by dashed lines back through the dichroic mirror, and focused with a suitable lens **1010** such as an f/1.4 camera lens on a linear detector **1012** via a variable f stop focusing lens **1014**. Through use of a linear light beam, it becomes possible to generate data over a line of pixels (such as about 1 cm) along the substrate, rather than from individual points on the substrate. In alternative embodiments, light is directed at a 2-dimensional area of the substrate and fluorescence is detected by a 2-dimensional CCD array. Linear detection is preferred because substantially higher power densities are obtained.

Detector **1012** detects the amount of fluorescence emitted from the substrate as a function of position. According to one embodiment the detector is a linear CCD array of the type commonly known to those of skill in the art. The x-y translation stage, the light source, and the detector **1012** are all operably connected to a computer **1016** such as an IBM PC-AT or equivalent for control of the device and data collection from the CCD array.

In operation, the substrate is appropriately positioned by the translation stage. The light source is then illuminated, and fluorescence intensity data are gathered with the computer via the detector.

In a preferred embodiment, the substrate and x/y translation table are placed under a microscope which includes one or more objectives. Light (488 nm) from a laser, which in some embodiments is a model no. 2020-05 argon ion laser manufactured by Spectra Physics, is directed at the substrate by a dichroic mirror which passes greater than about 520 nm light but reflects 488 nm light. The dichroic mirror may be, for example, a model no. FT510 manufactured by Carl Zeiss. Light reflected from the mirror then enters the microscope which may be, for example, a model no. Axioskop 20 manufactured by Carl Zeiss. Fluorescein-marked materials on the substrate will fluoresce >520 nm light, and the fluoresced light will be collected by the microscope and passed through the mirror. The fluorescent light from the substrate is then directed through a wavelength filter and, thereafter through an aperture plate. The wavelength filter may be, for example, a model no. OG530 manufactured by Melles Griot and the aperture plate may be, for example, a model no. 477352/477380 manufactured by Carl Zeiss.

The fluoresced light then enters a photomultiplier tube which in some embodiments is a model no. R943-02 manufactured by Hamamatsu, the signal is amplified in a preamplifier and photons are counted by a photon counter. The

60

number of photons is recorded as a function of the location in the computer. The pre-amp may be, for example, a model no. SR445 manufactured by Stanford Research Systems and the photon counter may be a model no. SR430 manufactured by Stanford Research Systems. The substrate is then moved to a subsequent location and the process is repeated. In preferred embodiments the data are acquired every 1 to 100 μm with a data collection diameter of about 0.8 to 10 μm preferred. In embodiments with sufficiently high fluorescence, a CCD detector with broadfield illumination is utilized.

FIG. **36** illustrates the architecture of the data collection system in greater detail. Operation of the system occurs under the direction of the photon counting program **1102**. The user inputs the scan dimensions, the number of pixels or data points in a region, and the scan speed to the counting program. Via a GPIB bus **1104** the program (in an IBM PC compatible computer, for example) interfaces with a multi-channel scaler **1106** such as a Stanford Research SR 430 and an x-y stage controller **1108** such as a Newport PM500. The signal from the light from the fluorescing substrate enters a photomultiplier **1110**, providing output to the scaler **1106**. Data are output from the scaler indicative of the number of counts in a given region. After scanning a selected area, the stage controller is activated with commands for acceleration and velocity, which in turn drives the scan stage **1112** such as a Newport PM500-A to another region.

Data are collected in an image data file **1114** and processed in a scaling program **1116**. A scaled image is output for display on, for example, a VGA display **1118**. The image is scaled based on an input of the percentage of pixels to clip and the minimum and maximum pixel levels to be viewed. The system outputs for use the min and max pixel levels in the raw data.

B. Data Analysis

The output from the data collection system is an array of data indicative of fluorescence intensity versus location on the substrate. The data are typically taken over regions substantially smaller than the area in which synthesis of a given polymer has taken place. Merely by way of example, if polymers were synthesized in squares on the substrate having dimensions of 500 microns by 500 microns, the data may be taken over regions having dimensions of 5 microns by 5 microns. In most preferred embodiments, the regions over which fluorescence data are taken across the substrate are less than about $\frac{1}{2}$ the area of the regions in which individual polymers are synthesized, preferably less than $\frac{1}{10}$ the area in which a single polymer is synthesized, and most preferably less than $\frac{1}{100}$ the area in which a single polymer is synthesized. Hence, within any area in which a given polymer has been synthesized, a large number of fluorescence data points are collected.

A plot of the number of pixels versus fluorescence intensity for a scan of a cell when it has been exposed to, for example, a labeled antibody will typically take the form of a bell curve, but spurious data are observed, particularly at higher intensities. Since it is desirable to use an average of fluorescence intensity over a given synthesis region in determining relative binding affinity, these spurious data will tend to undesirably skew the data.

Accordingly, in one embodiment of the invention the data are corrected for removal of these spurious data points, and an average of the data points is thereafter utilized in determining relative binding efficiency.

FIG. **37** illustrates one embodiment of a system for removal of spurious data from a set of fluorescence data such as data used in affinity screening studies. A user or the

US 6,420,169 B1

61

system inputs data relating to the chip location and cell corners at step 1302. From this information and the image file, the system creates a computer representation of a histogram at step 1304, the histogram (at least in the form of a computer file) plotting number of data pixels versus intensity.

For each cell, a main data analysis loop is then performed. For each cell, at step 1306, the system calculates the total fluorescence intensity or number of pixels for the bandwidth centered around varying intensity levels. For example, as shown in the plot to the right of step 1306, the system calculates the number of pixels within the band of width w . The system then "moves" this bandwidth to a higher center intensity, and again calculates the number of pixels in the bandwidth. This process is repeated until the entire range of intensities have been scanned, and at step 1308 the system determines which band has the highest total number of pixels. The data within this bandwidth are used for further analysis. Assuming the bandwidth is selected to be reasonably small, this procedure will have the effect of eliminating spurious data located at the higher intensity levels. The system then repeats at step 1310 if all cells have been evaluated, or repeats for the next cell.

At step 1312 the system then integrates the data within the bandwidth for each of the selected cells, sorts the data at step 1314 using the synthesis procedure file, and displays the data to a user on, for example, a video display or a printer.

C. Alternative Embodiments

Alternative embodiments of the detection system will be used according to some embodiments of the invention. According to one embodiment of the invention, a slit scanning fluorescence detection system is used in imaging VLSIPS™ chips. Such systems may have improved sensitivity, resolution, contrast, speed of data acquisition, etc. as compared to a pinhole system. Such systems have improved speed of data acquisition, since the image of the VLSIPS™ chip will be constructed of strips rather than scan lines. High resolution scans of VLSIPS™ chips can take over an hour to acquire with a pinhole system. The advantages of the slit scanning approach depends on the size of the imaged illumination slit (limited in practice by the size of the detector and the magnification of the optical system) and the sensitivity and dark noise of the linear detector used (i.e. how fast the detector is).

Better optical sectioning and hence reduction of background have been reported for a beam-scanning version of this approach. Theoretical calculations and experimental measurements for a scanning mirror/slit microscope using a divided aperture indicate that the contribution to image formation by out-of-plane light scattering/emitting elements will fall off faster with distance than in the case of a pinhole aperture system (Koester (1989), in *Handbook of Biological Confocal Microscopy*, pp 207–214, edited by J. Pawley, Plenum Press, N.Y., incorporated herein by reference for all purposes). Beam scanning arrangements are unlikely to be of use in near future because of the limited field of view obtained.

Slit scanning systems have been built by others (see references in *Handbook of Biological Confocal Microscopy*, Chapters 1 and 19, incorporated herein by reference).

Other improvements may be made to the system described elsewhere herein. For example, improved image contrast may be obtained by using dielectric barrier filters and a higher numerical aperture microscope objective. These combined modifications improve the detection limit by reducing background fluorescence and laser light scattering. A difficulty in using higher numerical aperture objec-

62

tives is the shallow depth-of-field, which leads to out-of-focus scans if the thickness of the substrates is nonuniform. As a first step towards solving this problem, a multiple focus and extrapolated x-axis control mechanism may be utilized.

An alternative solution to this problem involves the use of a piezoelectric focusing system which moves the monitor focus (e.g. by bouncing a focused laser off the surface and detecting positional variation with a diode array), a piezoelectric control device may be used to provide real-time autofocus capability.

Alternate fluorophores would also be beneficial in some embodiments. Fluorophores should be evaluated in terms of relative quantum yield, photobleaching stability, and detection sensitivity achieved under the scan conditions we use. To optimize detection sensitivity, fluorophores with larger Stokes' shifts allow better discrimination between emitted light and scattered laser light. In addition, they are less subject to self-quenching phenomena at high packing densities and hence should provide better quantitation of the relative number of fluorophores bound to the surface. Fluorophores with other excitation parameters may also be desirable in some embodiments. For example, Pharmingen makes available of a fluorophore which can be excited with the argon-ion laser and emits above 670 nm.

According to the work of Hirschfeld (*Appl. Optics* (1976) 15:3135–3139, incorporated herein by reference), the integrated fluorescent emission obtained upon complete bleaching of a fluorescent tag is independent of fluorescence quantum efficiency, absorption cross-section, and illumination intensity. Hence, this approach offers high sensitivity and better quantitation (a greater number of fluorophores can be attached to an antibody or packed into a small area without loss of signal due to quenching), and measurements should be less sensitive to errors in focusing than the scanning approach. The method is probably most suitable for sequential sampling of a small number of sites on a VLSIPS™ surface, although it should also be possible to build a device with a two-dimensional detector for simultaneous readout of many sites, provided that good rejection of excitation light and background fluorescence is achieved.

Time-resolved fluorescence provides an additional approach to enhanced sensitivity through background reduction. Background fluorescence in biological samples usually decays on the time scale of nano- to microseconds. Pulsed excitation of a fluorescent tag having a long lifetime can be detected with high sensitivity by gating the detector so that emitted light is measured after the background has decayed. Immunoassays have been developed using this approach with sensitivities that are reported to approach that of radioisotopic methods (Soini and Kojola (1983) *Clin. Chem.* (1983) 29:65–68, incorporated herein by reference). This mode of detection is particularly attractive for two reasons. First, considerable experience with rare-earth chelates having long fluorescence lifetimes is available (on the order of 1 msec), and may be able to provide novel additional compounds. Second, this approach allows one to image the VLSIPS™ surface using a two-dimensional detector (e.g., on CCD), resulting in reduced data acquisition time and an instrument that is likely to be much easier to utilize. The use of two-dimensional detectors in our present system is precluded by the necessity of using a pinhole aperture in front of the detector to achieve the confocal condition necessary for reduced background. For example, terbium and europium chelates are typically excited in the UV (320–340 nm), and emit at much longer wavelengths (545 nm for terbium and 615 nm for europium). Thus, they offer excellent wavelength discrimination as well as time-resolved

US 6,420,169 B1

63

discrimination against background fluorescence. To utilize these compounds a pulsed excitation source is utilized (e.g. an acousto-optically modulated argon-ion laser, flash lamp, or pulsed laser), a gated detector, and timing instrumentation.

Chemiluminescence has been reported to provide detection sensitivities comparable to that achieved using radioisotopes, and several products that can be chemically or enzymatically triggered to emit light are commercially available. High detection sensitivity results from measurement of signals against "zero" background (i.e. there's no background excitation light). In preliminary measurements using a commercially-available chemiluminescent substrate for alkaline phosphatase (Lumi-Phos 530; Boehringer-Mannheim #1275-470), it was possible to detect approximately 10 pmoles of enzyme in a 600 microliter sample using the detection end of the Aminco spectrofluorimeter. The lifetime of the unstable dioxetane intermediate in those experiments appeared to be too long to permit useful imaging of a VLSIPS™ surface in some embodiments (diffusion of the substrate would create a resolution problem if one used immobilized enzyme), but other compounds may have shorter lifetimes. Alternatively, using tethered substrate (one photon maximum released per surface site) may provide enough light if the collection efficiency were extremely high. In some embodiments, sandwiching a chemiluminescent probe-labeled VLSIPS™ chip between a CCD detector and a mirror to maximize collection efficiency may be utilized, while a fiber optic faceplate between chip and detector may be utilized to minimize cross-talk.

Other approaches may be utilized in detection of receptor-ligand interactions in some embodiments, e.g. ChemFETS. ChemFETS are semiconductor devices in which the current flowing through the device is modulated by electrostatic interactions between ions in solution and a region of the surface called the "gate". A multiple gate device may be utilized in some embodiments in which different macromolecules (e.g. receptors or antibodies) are immobilized at each gate. Detection of ligand binding may then be possible, either directly for charged ligands, or by using an enzyme-antibody conjugate that gives rise to local pH changes, by monitoring current for each gate region. A related device has recently been commercialized by Molecular Devices.

V. OTHER REPRESENTATIVE APPLICATIONS

A. Oligonucleotide Synthesis

The generality of light directed spatially addressable parallel chemical synthesis is demonstrated by application to nucleic acid synthesis.

1. EXAMPLE

Light activated formation of a thymidine-cytidine dimer was carried out. A three dimensional representation of a fluorescence scan showing a 7 square by 4 square checkerboard pattern generated by the light-directed synthesis of a dinucleotide was produced. 5'-nitroveratryl thymidine was attached to a synthesis substrate through the 3'hydroxyl group. The nitroveratryl protecting groups were removed by illumination through a 500 μ m checkerboard mask. The substrate was then treated with phosphoramidite activated 2'-deoxycytidine. In order to follow the reaction fluorometrically, the deoxycytidine had been modified with an Fmoc protected aminohexyl linker attached to the exocyclic amine (5'-O-dimethoxytrityl-4-N-(6-N-fluorenylmethylcarbamoyl-hexylcarboxy)-2'-deoxycytidine). After removal of the Fmoc protecting group with base, the regions which contained the dinucle-

64

otide were fluorescently labelled by treatment of the substrate with 1 mM FITC in DMF for one hour.

The three-dimensional representation of the fluorescence intensity data showing alternating squares of bright raised pixels reproduces the checkerboard illumination pattern used during photolysis of the substrate. This result demonstrates that oligonucleotides as well as peptides can be synthesized by the light-directed method.

2. EXAMPLE

In another example the light-activated formation of thymidine-cytidine was carried out as shown in FIG. 38. Here, as in the previous example, 5'-nitroveratryl thymidine was attached to the substrate, via phosphoramidite chemistry to a surface containing [Bis (2-hydroxyethyl)-3-aminopropylsiloxane]. The slide was then uniformly illuminated (362 nm at ~14 mW/cm²) for 10 minutes in the presence of dioxane. After drying, the surface was then treated with N,4-dimethoxytrityl-5'-nitroveratryl-2'-deoxycytidine-3'-O-(2-cyanoethyl)-N,N-diisopropylphosphoramidite in the presence of tetrazole (standard phosphoramidite coupling chemistry). After oxidizing and drying, the plate was again illuminated as before except that a 500 μ m checkerboard mask was placed between the light source and the slide. The surface was then exposed to 5'-O-(4,4'-Dimethoxy)-N-4-(6-((Biotinoyl)amino)hexanoyl)amino)hexanoyl, aminohexyl)-5-methyl-2'-deoxycytidine-3'-O-(2-cyanoethyl)-N,N-diisopropylphosphoramidite with tetrazole. After oxidizing and drying, the areas which contained the trinucleotide were fluorescently labelled by treatment with FITC labeled streptavidin. A resulting representation of the fluorescence intensity data showed alternating bright and dark squares corresponding to the 500 μ m and checkerboard illumination pattern used during photolysis.

3. EXAMPLE

In another example, an 8 nucleotide, poly-adenine oligomer was prepared and later hybridized with a poly thymidine probe. The synthesis was carried out as follows.

Ten 1x3" slides were incubated in a plastic jar filled with 1% bis(2-hydroxyethyl)aminopropyltriethoxy-silane in 95% ethanol overnight at room temperature. The slides were then rinsed thoroughly with ethanol, dried with N₂ and baked at 110° C. for 1 hour and put in a vacuum desiccator to cool. A surface linker for coupling was prepared by mixing equal volumes of 0.2M monodimethoxytritylpentaethyleneglycol- β -cyanoethylphosphoramidite in anhydrous acetonitrile and 0.45M tetrazole/CH₃CN in a glass vial. 0.35 mL of this solution was then dispensed onto the surface of each slide and incubated for 3 min. The slide was rinsed briefly with CH₃CN and coupling was repeated with freshly prepared phosphoramidite. Next, the phosphite-triester bond was oxidized to a phosphotriester by dipping the slides into a jar filled with 0.1M iodine solution (2.6 g iodine+80 mL tetrahydrofuran+20 mL 2,6-lutidine+2 mL H₂O) for 1 min, followed by rinsing thoroughly with CH₃CN and drying with N₂. The dimethoxytrityl protecting groups were removed by dipping the slides in a staining jar filled with 3% dichloroacetic acid in methylene chloride for 30 sec followed by rinsing with CH₃CH and drying with N₂. Steps C and D were then repeated with 0.2M 5'-nitroveratryl deoxythymidine-3'- β -cyanoethylphosphoramidite. The slides were then incubated for 1 hour in a staining jar filled with capping solution (75 mLs 6.5% DMAP/THF+25 mLs 40% acetic anhydride/60% 2,6-lutidine), and then rinsed

US 6,420,169 B1

65

thoroughly with CH_3CN , dried with N_2 , and stored in the dark under vacuum.

Polydeoxyadenine (poly-dA₁₂) was prepared on an ABI synthesizer using 1 μmole 3'-amine-ON CPG (Glen Research) and the standard ABI 1 μmole coupling cycle having final DMT on. The CPG was transferred to a 1.5 mL plastic screw-cap vial where 1.0 mL conc. NH_4OH was added. The mixture was incubated for 18 hours at 55° C. to cleave the oligomer from the resin and remove the exocyclic amine protecting groups. The crude oligomer was purified using a PolyPak cartridge (Glen Research) according to the protocol supplied with the columns. The appropriate fractions were pooled and dried by speed-vac. The oligomer was dissolved in 0.9 mL H_2O , to which 0.1 mL 10 \times labelling buffer (1.0M $\text{NaHCO}_3/\text{Na}_2\text{CO}_3$, pH 9.0) and 0.25 mL freshly prepared 100 mg/mL FITC in DMF were added. The solution was then vortexed and incubated overnight at room temperature. The reaction mixture was purified on a 1 \times 30 cm Sephadex G-25™ column using H_2O as the mobile phase, and the appropriate fractions were pooled and dried by speed-vac. PAGE analysis showed that the reaction was not 100% complete. The fluoresceinylated oligomer was further purified by reverse phase HPLC (Hamilton PRP-1 semi-preparative column) using a linear gradient of 10–40% CH_3CN in 0.1M Triethylamine acetate, pH 7.6 over 45 min. at a flow rate of 2 mL/min. Fractions were analyzed by PAGE, and the appropriate fractions were pooled and dried by speed-vac. PAGE analysis of the final product showed purity of approximately 99%.

A suitably derivatized slide was placed in a four-chamber flow cell (wells approximately 1.5 cm dia. circles). One well was filled with dioxane. All other wells were covered with black electrical tape. The slide was then exposed to 365 nm light at 11.8 mW/cm² for a period of 12 min. to remove the photoprotecting groups. The flow cell was then attached to an ABI DNA synthesizer (model 392), and 5'-NV-dT-OCEP coupled to the surface using the modified cycle attached (cyc03 user). No capping step was performed due to the excess phosphoramidite used. The photolysis/synthesis cycles were repeated until poly-(dT)₁₂-OH was synthesized in the well. The slide was removed from the flow cell, rinsed thoroughly with CH_3CN and dried with N_2 . The slide was then incubated in 6 \times SSPE containing 3% BSA and 0.025% triton X-100 for 30 min. at room temperature to block non-specific binding sites. Next, the slide was transferred to a plastic container filled with 20 ng/mL 5'-HO-poly(dA)₁₂-fluorescein in the same buffer and incubated at room temperature for 1 hour. The slide was briefly rinsed in a 20 mM NaCl solution, dried with N_2 , and detected using the confocal microscopy system previously described. Average photon counts in the well were 8-fold higher than the background producing the image shown in FIG. 39A. The slide was incubated overnight in 500 mLs 1 \times SSPE at 40° C., rinsed and scanned. FIG. 39A shows a bright circle in the center indicating that this wash removed the signal as shown in FIG. 39B. Reprobing as before resulted in a signal with the same intensity as that originally obtained as shown in FIG. 39C.

For sample 59-8b-1, -2, and -3, the average intensities were 347 \pm 25, 211 \pm 23 and 223 \pm 13, respectively, in the background. For samples 59-8b-1 and 59-8b-3 the signal in the well was 4545 \pm 476 and 237 \pm 308, respectively. Therefore, for sample 1 the signal/background ratio was 13, and it was 11 for sample number 3. The sample number 59-8b-1 refers to the first probe. The sample number 59-8b-2 refers to the scan after incubation at 40° C. The sample 59-8b-3 is the reprobe sample.

66

B. Oligosaccharide Synthesis

The present invention will find application in a wide variety of additional applications including oligosaccharide synthesis.

1. EXAMPLE

The potential synthesis of a C-glycoside unit that will be attached to the surface is outlined below. FIG. 40A illustrates the overall synthesis strategy.

The methyl glycoside of D-glucose is first converted through a standard protection-deprotection sequence to the diacetate. Selective esterification of the primary alcohol is then conducted, followed by etherification of the secondary alcohol with a photolabile protecting group and C-allylation, which provides the C-glycoside. Oxidation of the olefin effects conversion to the carboxylic acid which is then attached to the surface through standard amide coupling. Photodeprotection will free a hydroxyl group on the monosaccharide for the purposes of examining glycosidic bond formation on the chip surface.

FIG. 40B illustrates formation of a simple activated building block used especially for purposes of examining chemical glycosidic bond formation. The building block is formed D-glucose.

Conversion to the 6-amino derivative is effected through fairly simple methods, such as shown in the first step of the Figure. Protection of the amino function followed by exhaustive acetylation provides the fully protected methyl glycoside. Conversion to an activated glycosyl donor is done by hydrolysis and conversion to the bromide. This material may be attached to a substrate as shown in FIG. 40C. Coupling of the activated bromide to the surface attached glycosyl acceptor will be monitored through the agency of protected 6-amino function on the donor. Photodeprotection followed by labeling with fluorescein isothiocyanate will provide a sensitive assay for the formation of the glycosidic bond.

Examination of enzymatic glycosidic bond formation requires the synthesis of a nucleoside diphosphate sugar. One possible scheme is illustrated in FIG. 40D. The introduction of the amino group and exhaustive acetylation is accomplished by standard methods such as those shown in the Figure. Protection of the amine and glycoside hydrolysis leads to the free sugar which is now activated via conversion to the bromide followed by displacement to the anomeric phosphate derivative. Conversion to the uridine diphosphate (UDP) derivative is accomplished through fairly standard methodology, such as the one shown. The product may then be attached to the substrate as shown in FIG. 40E.

Formation of the glycosidic bond between the nucleotide diphosphate sugar and the immobilized saccharide on the VLSIPS™ chip with galactosyl transferase may also be examined. Coupling of the nucleoside glycoside donor to the surface attached glycosyl acceptor will be monitored through the agency of the 6-amino function on the donor. Labeling of the free amine with fluorescein isothiocyanate will provide a sensitive assay for the formation of the glycosidic bond.

C. Caged Binding Member System

According to an alternative embodiment of the invention, the methods provide for attaching to the surface a caged binding member which in its caged form has a relatively low affinity for other potentially binding species, such as receptors and specific binding substances. Such techniques are more fully described in copending application U.S. Ser. No. 404,920, filed Sep. 8, 1989, and incorporated herein by reference for all purposes.

US 6,420,169 B1

67

According to this alternative embodiment, the invention provides methods for forming predefined regions on a surface of a solid support, wherein the predefined regions are capable of immobilizing receptors. The methods make use of caged binding members attached to the surface to enable selective activation of the predefined regions. The caged binding members are liberated to act as binding members ultimately capable of binding receptors upon selective activation of the predefined regions. The activated binding members are then used to immobilize specific molecules such as receptors on the predefined region of the surface. The above procedure is repeated at the same or different sites on the surface so as to provide a surface prepared with a plurality of regions on the surface containing, for example, the same or different receptors. When receptors immobilized in this way have a differential affinity for one or more ligands, screenings and assays for the ligands can be conducted in the regions of the surface containing the receptors.

The alternative embodiment may make use of novel caged binding members attached to the substrate. Caged (unactivated) members have a relatively low affinity for receptors of substances that specifically bind to uncaged binding members when compared with the corresponding affinities of activated binding members. Thus, the binding members are protected from reaction until a suitable source of energy is applied to the regions of the surface desired to be activated. Upon application of a suitable energy source, the caging groups labilize, thereby presenting the activated binding member. A typical energy source will be light.

Once the binding members on the surface are activated they may be attached to a receptor. The receptor chosen may be a monoclonal antibody, a nucleic acid sequence, a drug receptor, etc. The receptor will usually, though not always, be prepared so as to permit attaching it, directly or indirectly, to a binding member. For example, a specific binding substance having a strong binding affinity for the binding member and a strong affinity for the receptor or a conjugate of the receptor may be used to act as a bridge between binding members and receptors if desired. The method uses a receptor prepared such that the receptor retains its activity toward a particular ligand.

Preferably, the caged binding member attached to the solid substrate will be a photoactivatable biotin complex, i.e., a biotin molecule that has been chemically modified with photoactivatable protecting groups so that it has a significantly reduced binding affinity for avidin or avidin analogs than does natural biotin. In a preferred embodiment, the protecting groups localized in a predefined region of the surface will be removed upon application of a suitable source of radiation to give binding members, that are biotin or a functionally analogous compound having substantially the same binding affinity for avidin or avidin analogs as does biotin.

In another preferred embodiment, avidin or an avidin analog is incubated with activated binding members on the surface until the avidin binds strongly to the binding members. The avidin so immobilized on predefined regions of the surface can then be incubated with a desired receptor or conjugate of a desired receptor. The receptor will preferably be biotinylated, e.g., a biotinylated antibody, when avidin is immobilized on the predefined regions of the surface. Alternatively, a preferred embodiment will present an avidin/biotinylated receptor complex, which has been previously prepared, to activated binding members on the surface.

D. Fingerprinting for Quality Control

An alternative aspect of this invention involves testing a therapeutic compound with an array of peptides or other

68

biological polymers to determine a characteristic binding pattern. Such a characteristic binding pattern or "fingerprint" is used to monitor the "constancy" of the compound over time by repeated testing with the same array. As long as the fingerprint remains unchanged from lot-to-lot, the bioprocess is producing the same product. If the binding pattern of the therapeutic changes at any time, it would be assumed that a subtle (or not so subtle) change in the therapeutic compound had occurred. For example, changes in the glycosylation or secondary structure of the protein could be detected. This method promises to be particularly valuable with recombinant and other products produced by fermentation processes where quality control is problematic.

Preferably, the method would be performed with a very large array of biological polymers (thousands or tens of thousands of elements). A VLSIPS™ or caged biotin chip with binding components to a very broad spectrum of polymer characteristics would be employed. The elements of the array could thus be arranged in a quasi-random manner. Alternatively, they might be organized in a rational order. For example, certain physical properties (e.g., charge or hydrophobicity) of the constituent polymers might vary gradually along a given spatial dimension. Chips having this configuration could then be used for the QC of various biological products. Custom chips specific for a single product, such as for example tissue plasminogen activator (tPA), might also be useful.

E. β -Amino Acid and D-Amino Acid Monomers

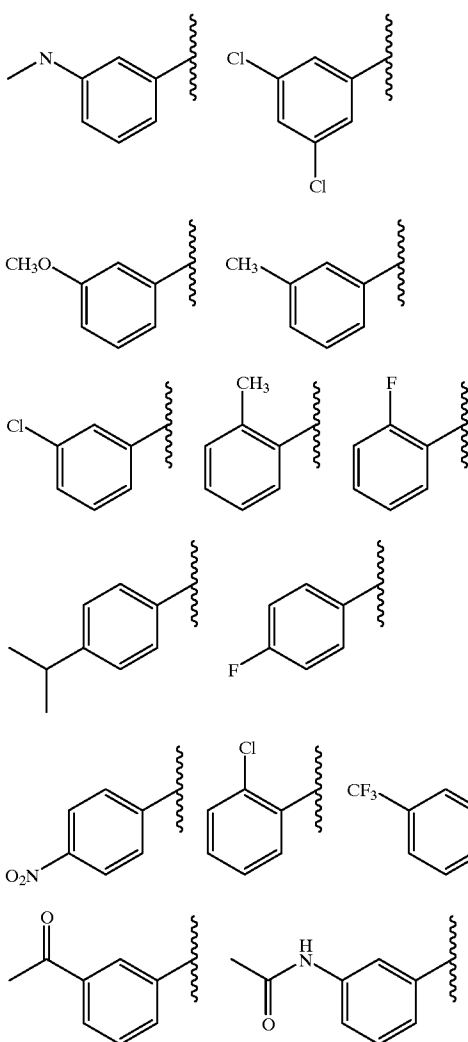
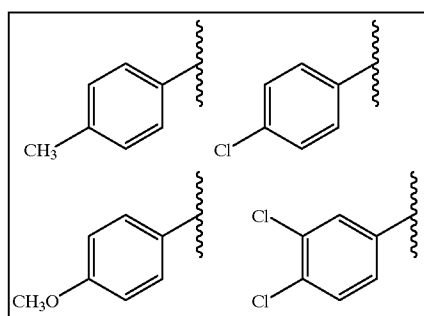
The peptide diversity available in the present invention is greatly increased by including non-natural amino acids (i.e. amino acids which are not genetically coded) among the set of building blocks. In particular, peptides containing at least one β -amino acids or D-amino acid residue may be synthesized by the methods of the present invention. In addition, L-amino or D-amino acids having modified side chains may be employed to diversify the peptide products available. Cyclic β -amino acid monomers may be employed to reduce the increased conformational mobility associated with acyclic β -amino acids. In fact, an ordered series of dihedral angles between the amido and carboxamido groups of the peptide backbone may be obtained by changing the number and constituent atoms of the cyclic β -amino acid ring. Such an ordered series of compounds may be useful in optimizing the biological activity of a peptide drug. Since D-amino acids and β -amino acids of any sort are not subject to proteolytic cleavage, incorporation of these residues into peptide drugs should confer favorable properties of biological stability.

Amino acid monomers (D-, L- or β -) with side chains that are not found on the genetically coded amino acids may also be used in preparing peptides according to the present invention. Amino acids containing aromatic residues are of particular interest because they are commonly present in biologically active binding sites and in drugs. Although they are generally hydrophobic, aromatic side chains can adopt a variety of electronic configurations depending upon the substituents present.

Using phenylalanine as an example, a variety of modifications to the side chain are available, some of which are represented below. Each amino acid containing the groups below can be employed as a monomer without requiring side chain protecting groups often necessary for peptide synthesis.

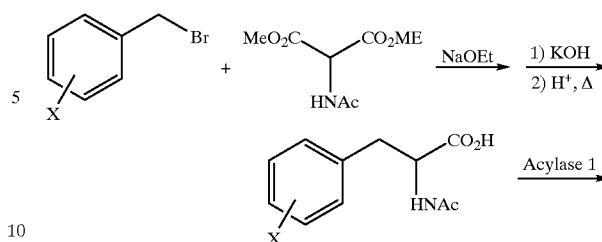
US 6,420,169 B1

69



Many of these compounds are commercially available. Those that are not may be synthesized by a variety of well-known procedures. A preferred choice is alkylation of acetamidomalonate with the appropriate benzylic halide, followed by hydrolysis, decarboxylation and enzymatic resolution with *Aspergillus acylase* I as shown below and described in Chenault et al., *J. Am. Chem. Soc.* (1989) 111:6354-6364 which is incorporated herein by reference for all purposes.

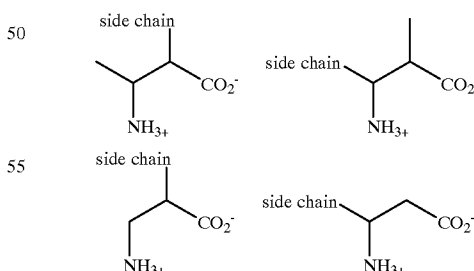
70



Like genetically coded L-amino acids, D-amino acids must be protected at the α -amino group during peptide synthesis. In addition, the side chain may also have to be protected against unwanted side reactions. The methods of protection set forth above for L-amino acids can also be applied to D-isomers and amino acids with modified side chains. The resulting protected monomers can then be employed to synthesize an array of peptides (or peptide analogs) using the protection-deprotection methods outlined above.

Polymer backbones comprised of β -amino acids have several advantages. For example, they retain amide bonds, which permits hydrogen bonding in all directions normal to the main chain. The side chain density will be high for a given oligomer length so long as the side chains are placed at both α and β carbons. Considerable control over the properties of the peptide is gained by selecting appropriate substituents. For example, attaching an alkyl group to the β carbon, makes the backbone more hydrophobic than the corresponding α -amino acids. Additional control of the peptide conformation is also possible. Although extra conformational freedom is permitted by the β carbon in straight chain β -amino acids, conformational restriction within the individual monomer units is also possible by selecting appropriate side chains and cyclic groups, as will be shown below.

Simple β -amino acids which may be used in the present invention will have an amino acid side chain at either the α or β carbon, and a methyl group or hydrogen atom at the other (types I through IV shown below). Each of the methyl compounds (types I and II below) will include four distinct stereoisomers for each amino acid chain, representing a total of 160 compounds when all of the genetically-coded side chains are employed. Replacing the methyl group with a hydrogen atom (groups III and IV), reduces the number of stereoisomers to two for each side chain.



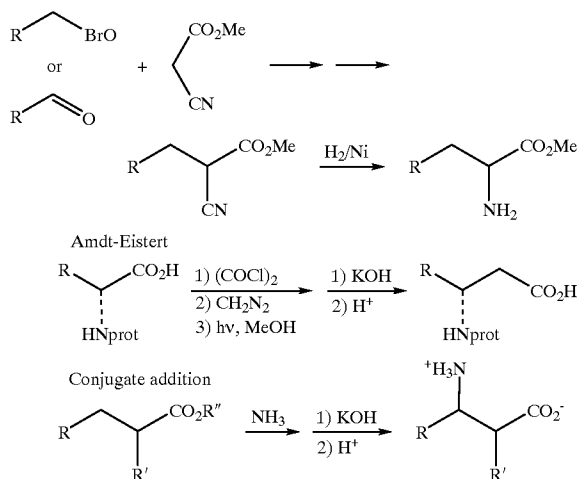
Three methods for preparing β -amino acids shown below are summarized in Griffith, *Annu. Rev. Biochem.* (1989) 55:855-878 which is incorporated herein by reference for all purposes. The condensation of cyanoacetic ester with carbonyl compounds or alkyl halides followed by reduction provides structures of type III. The Arndt-Eistert homolog-

US 6,420,169 B1

71

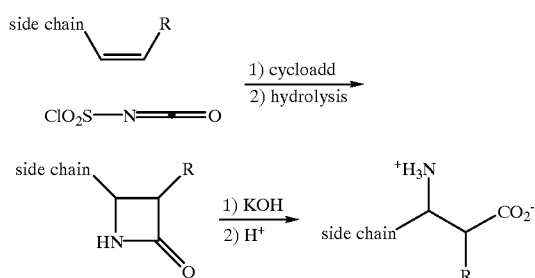
gation of protected amino acids will give compounds having the structures of type

IV. CONJUGATE ADDITION OF AMMONIA TO α , β -UNSATURATED ESTERS WILL PRODUCE COMPOUNDS OF TYPES I-IV, DEPENDING ON SUBSTITUENTS IN THE ESTER STARTING MATERIAL



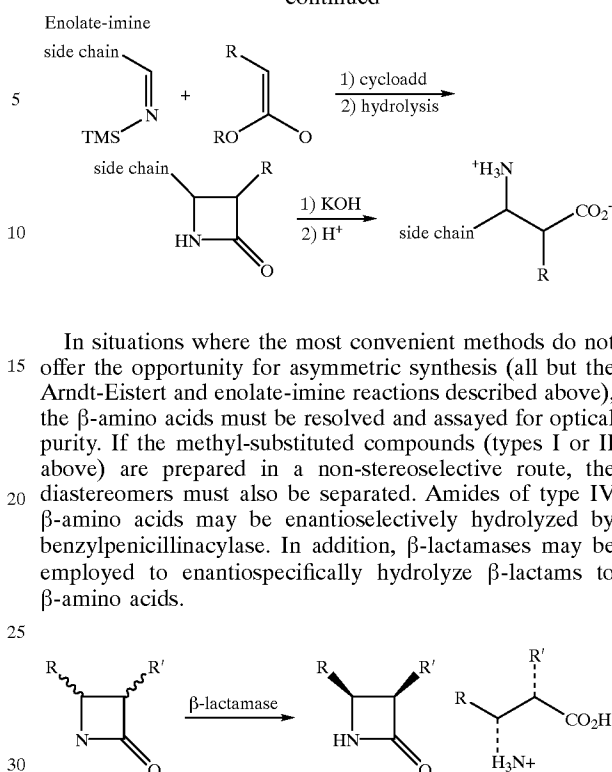
Two other synthetic routes (shown below) which may be employed to synthesize β -amino acids take advantage of the functional relation of β -lactams (which have well-known chemistries) and β -amino acids. These methods are detailed in various references, including Kamal et al., *Heterocycles* (1987) 26:1051-1076 and Hart et al., *Chem. Rev.* (1989) 89:1447-1465, both of which are incorporated herein by reference for all purposes. The first synthetic route exploits the cycloaddition reactions of chlorosulfonyl isocyanate (CSI) with alkenes to give, after hydrolysis, β -lactams. These can then be hydrolyzed to give the corresponding the β -amino acids. Because of the polar mechanism for the CSI cycloaddition, it is not possible to use this reaction to prepare type III compounds which have no substitution on the carbon atom adjacent to the nitrogen atom. The second route employs the condensation of enolates with imines, to produce β -lactams. Optically active compounds are provided this method, but the basic reaction with diazomethane gives only type IV structures. Type II structures may be prepared if diazomethane is substituted with diazoethane and the resulting diastereomers are separated. This synthesis route has the added advantage that it may directly provide protected amino acids for peptide synthesis.

Chlorosulfonylisocyanate

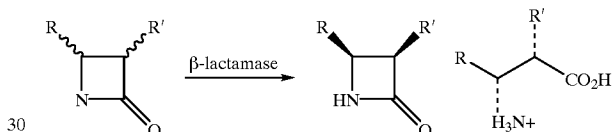


72

-continued



In situations where the most convenient methods do not offer the opportunity for asymmetric synthesis (all but the Arndt-Eistert and enolate-imine reactions described above), the β -amino acids must be resolved and assayed for optical purity. If the methyl-substituted compounds (types I or II above) are prepared in a non-stereoselective route, the diastereomers must also be separated. Amides of type IV β -amino acids may be enantioselectively hydrolyzed by benzylpenicillinacylase. In addition, β -lactamases may be employed to enantiospecifically hydrolyze β -lactams to β -amino acids.



Other methods of separating the isomers will be known to those of skill in the art. In any event, analysis of the optical purity of the products can be accomplished by chiral chromatography or the Mosher method. The absolute configuration, which must be determined for each compound prepared, can be assigned by chiroptical methods.

One skilled in the art will be able to readily determine an appropriate strategy for synthesizing peptides from various β -amino acid monomers employed in the present invention. However, it should be noted that some classical methods of peptide coupling (mixed anhydride, DCC) will not work with some β -amino acids. While not wishing to be bound by theory, it is believed that the lack of a group at the α -carbon causes increased side reactions of the activated carboxyl group. Another possible side reaction which should be avoided is formation of a dihydrooxazinone, which has been observed in some cases and which may participate in the coupling reaction.

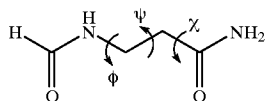
Successful synthesis strategies for some homopolymers (e.g., poly (β -amino butyrate)) include polymerization of β -lactams. See e.g. Chen et al., *Macromolecules* (1974) 7:779 which is incorporated herein by reference for all purposes. Similar methods may be employed in some instances with methods of the present invention. The homopolymers so produced have been found to adopt a β -conformation analogous to the poly(β -hydroxy butyrate) polymers produced by bacteria. The α -esters of aspartic acid also have been oligomerized to form β -peptides, and they form β -sheets when containing eight or more units. Directed synthesis in solution of a tripeptide formed from β -amino butyrate has been accomplished using trichlorophenyl active esters in the presence of hexamethylphosphoramide (HMPA) as described in Drey et al., *J. Chem. Soc., Perkin Trans.* (1982) 1:1587-1592 which is incorporated herein by reference for all purposes. Other syntheses well-known in

US 6,420,169 B1

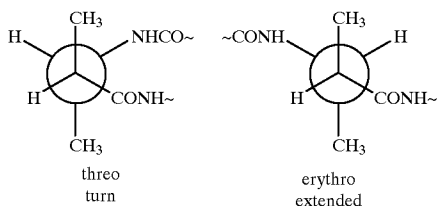
73

the art may be employed in the present invention to produce a variety of peptide oligomers.

Three dihedral angles (shown below) may be controlled in β -peptides.

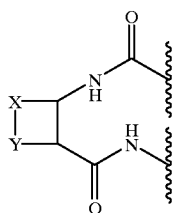


The angle depicted as Ψ above has no analogy in a α -peptides and is an angle over which there is appreciable control via the α and β substituents. Considering the erythro and threo isomers of α,β -dimethyl- β -alanine, a prototype disubstituted β -peptide, empirical force field calculations suggest that, in both cases, anti orientation of the methyl groups is favored. This results in the threo isomer introducing a turn in the chain, while the anti isomer would tend to maintain an extended backbone. Certainly, other conformations are close in energy, and will also be populated. However,



alkyl groups may be linked to a ring to control the rotation of the Ψ bond. This allows variation in the orientation of the main chain in a systematic way. A family of cyclic β -amino acids has been designed for this purpose. The size of the ring onto which the β -amino acid unit is fused and the fusion geometry limit the possibilities for the dihedral angle between the carboxy carbon and the amino group (angle Ψ). For example, the cis-cyclopropyl compound is constrained to eclipse these bonds ($\Psi=0^\circ$), while the trans-cyclopropyl locks them at a 144° angle. Energy minimization of other members of this homologous series showed an orderly progression of dihedral angles for several low energy conformers.

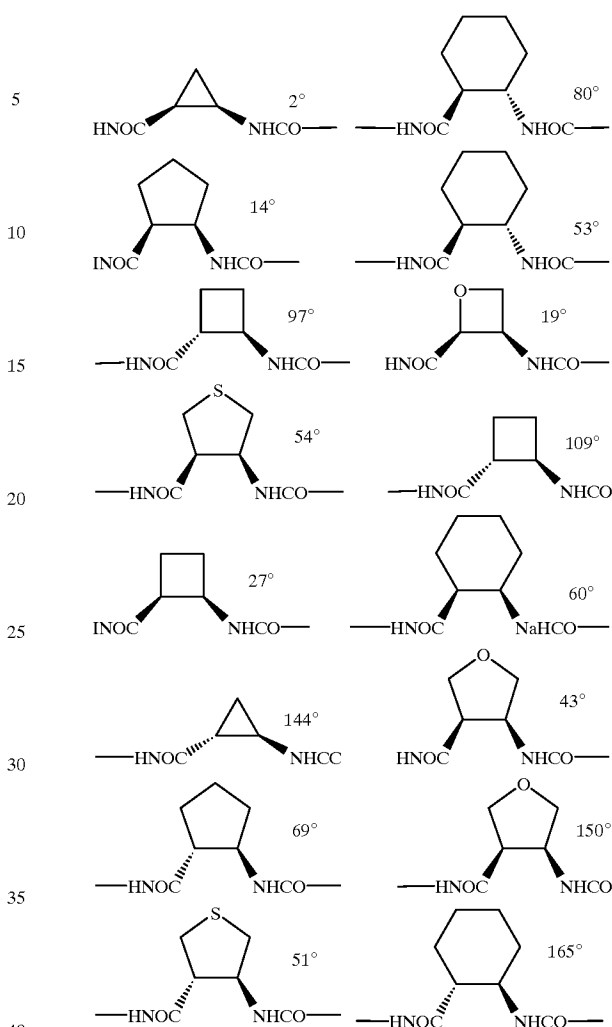
The generic structure for the cyclic β -amino acid monomers is shown below.



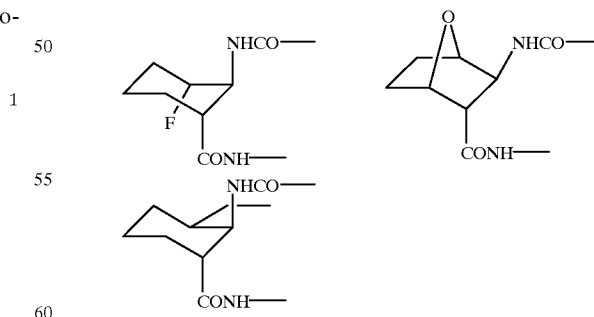
where X is carbon, silicon or the like, and Y is one or more carbon, nitrogen, oxygen, sulfur, silicon or no atom.

Depending on the enantiomeric series to be employed, the turn or extension introduced into the backbone can be either of positive or negative helicity.

74



Some of the compounds (e.g., the diaxial cyclohexane system) do not have as a global minimum in the conformation shown, and further constraints may be utilized to enforce the desired stereochemical relationship. These constraints will be well-known in the art and include, by way of example, polar effects, other rings, or allylic strain.

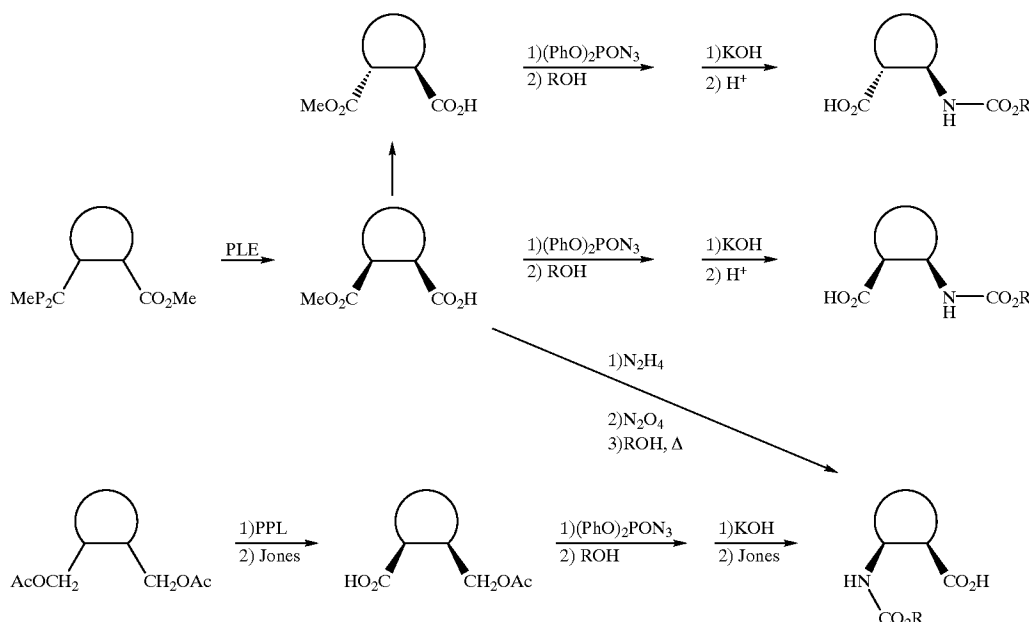


One advantage of this homologous series of compounds is that a unified synthetic approach can be employed. For example, the cycloaddition of chlorosulfonylisocyanate with cycloalkenes may be utilized. This is most applicable to cis isomers of cyclopentyl and cyclohexyl systems. Enantiomer selective enzymatic hydrolysis of the β -lactam then gives

US 6,420,169 B1

75

the desired β -amino acids ready for derivatization. A second route is more general, and longer. It begins with 1,2-cycloalkanedicarboxylic esters (or their derived diol acetates), which can be prepared via Diels-Alder reactions of maleates or from dianion alkylation of succinate as described in Garratt et al., *Tetrahedron Lett.* (1987) 28:351–352 which is incorporated herein by reference for all purposes. Enzymatic transformation then introduces optical activity and differentiates the carboxyl groups, which permits selective conversion of one into an amino group as described in Sabbioni et al., *J. Org. Chem.* (1987) 52:4565–4570 which is incorporated by reference herein for all purposes. One advantage of this strategy is that it directly provides the amino protected building block.



76

or other polymer. According to a first step of the process, a substrate is formed having an optional chain of amino acids (indicated by AA_n) and an exposed amino terminus. The substrate is reacted with an aldehyde derivative of a peptide in the presence of, for example, HCL or AcOH in DMF, providing a substrate with the chain of amino acids and a terminal imine group. The substrate is then reacted with, for example, NaBH₄ or NaCNBH₃ forming a sequence with a reduced amide bond in the growing polymer chain. The substrate is then optionally processed according to the methods described above to provide additional amino acids on the growing chain.

VI. CONCLUSION

The inventions herein provide a new approach for the simultaneous synthesis of a large number of compounds.

Other compounds related to the above β -amino acid analogues may be employed in peptide syntheses according to the present invention. For example, α -aminoxy acids in which the β -carbon of a β -amino acid is replaced with an oxygen may be used. These compounds are easily prepared in optically active form from the amino acids via the bromo acid. Several peptide analogues incorporating this unit have been synthesized in solution and are known to be resistant to mammalian proteases. See e.g. Briggs et al., *J. Chem. Soc. Perkin Trans.* (1979) 1:2138–2143 which is incorporated by reference herein for all purposes.

F. Reduced Amide Bonds

Reduced amide peptide isosteres have been incorporated into peptides by reductive alkylation to produce antagonists, enzyme inhibitors, and resistance to biodegradation. For example, in the case of a protease, one can render the scissile bond non-cleavable by introduction of the reduced amide at the cleavage site. In the case of Renin, a reduced peptide analog of the native Renin substrate (H-142) has been shown to reduce blood pressure in clinical tests.

According to some embodiments, the present invention provides for the introduction of a reduced amide bond into a growing peptide chain or other polymer chain on a substrate in situ, without the preformation of a dipeptide containing a reduced amide bond. FIG. 41 illustrates the introduction of a reduced amide bond into a growing peptide

The method can be applied whenever one has chemical building blocks that can be coupled in a solid-phase format, and when light can be used to generate a reactive group.

The above description is illustrative and not restrictive. Many variations of the invention will become apparent to those of skill in the art upon review of this disclosure. Merely by way of example, while the invention is illustrated primarily with regard to peptide, oligosaccharide and nucleotide synthesis, the invention is not so limited. By way of another example, while the detection apparatus has been illustrated primarily herein with regard to the detection of marked receptors, the invention will find application in other areas. For example, the detection apparatus disclosed herein could be used in the fields of catalysis, DNA or protein gel scanning, and the like. The scope of the invention should, therefore, be determined not with reference to the above description, but instead should be determined with reference to the appended claims along with their full scope of equivalents.

US 6,420,169 B1

77

78

 SEQUENCE LISTING

(1) GENERAL INFORMATION:

(iii) NUMBER OF SEQUENCES: 36

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 5 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Tyr Gly Gly Phe Leu
 1 5

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 5 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Pro Gly Gly Phe Leu
 1 5

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 6 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Tyr Pro Gly Gly Phe Leu
 1 5

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 6 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Tyr Gly Ala Phe Leu Ser
 1 5

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 5 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

US 6,420,169 B1

79

80

-continued

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Tyr Gly Ala Phe Ser
1 5

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Tyr Gly Ala Phe Leu
1 5

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Tyr Gly Gly Phe Leu Ser
1 5

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Tyr Gly Ala Phe
1

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Tyr Gly Ala Leu Ser
1 5

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single

US 6,420,169 B1

81

82

-continued

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Tyr Gly Gly Phe Ser
1 5

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 4 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Tyr Gly Ala Leu
1

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Tyr Gly Ala Phe Leu Phe
1 5

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 5 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Tyr Gly Ala Phe Phe
1 5

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 5 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Tyr Gly Gly Leu Ser
1 5

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 4 amino acids

(B) TYPE: amino acid

US 6,420,169 B1

83

84

-continued

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Gly Gly Phe Leu
1

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Tyr Gly Ala Phe Ser Phe
1 5

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Tyr Gly Ala Phe Leu Ser Phe
1 5

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Tyr Gly Ala Phe Met Gln
1 5

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Tyr Gly Ala Phe Met
1 5

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids

US 6,420,169 B1

85

86

-continued

(B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Tyr Gly Ala Phe Gln
 1 5

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 5 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Tyr Gly Gly Phe Met
 1 5

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 5 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Tyr Ala Gly Phe Leu
 1 5

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 5 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

Tyr Ser Gly Phe Leu
 1 5

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 5 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Tyr Pro Gly Phe Leu
 1 5

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

US 6,420,169 B1

87

88

-continued

- (A) LENGTH: 5 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

Leu Gly Gly Phe Leu
1 5

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 5 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

Phe Gly Gly Phe Leu
1 5

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 5 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

Leu Ala Gly Phe Leu
1 5

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 5 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

Phe Ala Gly Phe Leu
1 5

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 5 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

Trp Gly Gly Phe Leu
1 5

(2) INFORMATION FOR SEQ ID NO:30:

US 6,420,169 B1

89

90

-continued

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 5 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

Tyr Gly Ala Gly Phe
1 5

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 7 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

Arg Gln Phe Lys Val Val Thr
1 5

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 6 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

Arg Xaa Lys Val Val Thr
1 5

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 7 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

Arg Gln Xaa Lys Val Val Thr
1 5

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 8 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

Arg Gln Xaa Phe Lys Val Val Thr
1 5

(2) INFORMATION FOR SEQ ID NO:35:

US 6,420,169 B1

91

92

-continued

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 7 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

Arg Xaa Phe Lys Val Val Thr
 1 5

(2) INFORMATION FOR SEQ ID NO:36:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 4 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

Lys Val Val Thr

What is claimed is:

1. An apparatus for forming polynucleotides, said apparatus comprising:

a body having a cavity therein, said cavity being less than 1000 μm deep;

a substrate comprising a surface mated to said body, whereby said surface contacts and seals said cavity, wherein said surface has functional groups in fluid communication with said cavity and reactive with nucleosides or nucleotides, and

an inlet port and an outlet port, said inlet port and said outlet port being in fluid communication with said cavity.

2. An apparatus for forming polynucleotides, said apparatus comprising:

a body having a sealed cavity disposed therein, said cavity being less than 1000 μm deep and comprising a substrate having a surface, wherein said surface has functional groups reactive with nucleotides or nucleosides; and

an inlet port and an outlet port, said inlet port and said outlet port being in fluid communication with said cavity; and

a means for providing ultrasonic radiation in said cavity.

3. An apparatus for forming polynucleotides, said apparatus comprising:

a body having a sealed cavity disposed therein, said cavity being less than 500 μm deep;

a substrate comprising a surface mated to said body, whereby said surface contacts and seals said cavity, wherein said surface has functional groups in fluid communication with said cavity and reactive with nucleotides or nucleosides; and

an inlet port and an outlet port, said inlet port and said outlet port being in fluid communication with said cavity.

4. The apparatus as recited in claim 3 wherein said functional groups are protected with a protective group.

5. The apparatus as recited in claim 1 wherein the functional groups are protected with MeNPOC.

6. The apparatus of claim 4, wherein the protective group is nitroveratryloxycarbonyl.

7. An apparatus for forming polynucleotides, said apparatus comprising:

a glass substrate having a surface comprising linker molecules bearing functional groups reactive with nucleotides or nucleosides, said surface mated to a body having a cavity for sealing said cavity such that said linker molecules are in fluid communication with said cavity wherein said cavity is being less than 1000 μm deep;

an inlet port and an outlet port, said inlet port and said outlet port being in communication with said cavity, wherein said cavity is less than 1000 μm deep;

a fluid flowing means coupled to said inlet port.

8. An apparatus for forming polypeptides, said apparatus comprising:

a body having a sealed cavity disposed therein, said cavity being less than 1000 μm deep;

a substrate comprising a surface mated to said body, whereby said surface contacts and seals said cavity, wherein said surface has functional groups in fluid communication with said cavity and reactive with amino acids; and

an inlet port and an outlet port, said inlet port and said outlet port being in fluid communication with said cavity.

9. The apparatus of claim 8, further comprising a pump positioned to flow said fluid into said cavity through said inlet port and out of said cavity through said outlet port.

10. An apparatus for forming polypeptides, said apparatus comprising:

a body having a sealed cavity disposed therein, said cavity being less than 1000 μm deep and comprising a substrate having a surface, wherein said surface has functional groups reactive with amino acids; and

an inlet port and an outlet port, said inlet port and said outlet port being in fluid communication with said cavity; and

US 6,420,169 B1

93

a means for providing ultrasonic radiation in said cavity.

11. An apparatus for forming polypeptides, said apparatus comprising:

a body having a sealed cavity disposed therein, said cavity being less than 500 μm deep; 5

a substrate comprising a surface mated to said body, whereby said surface contacts and seals said cavity, wherein said surface has functional groups in fluid communication with said cavity and reactive with amino acids; and 10

an inlet port and an outlet port, said inlet port and said outlet port being in fluid communication with said cavity.

12. The apparatus as recited in claim **11** wherein said functional groups are protected with a protective group. 15

13. The apparatus as recited in claim **12** wherein said protective group comprises MeNPOC.

14. The apparatus of claim **12**, wherein the protective group is nitroveratryloxycarbonyl.

94

15. An apparatus for forming polypeptides, said apparatus comprising:

a glass substrate having a surface comprising linker molecules with functional groups reactive with amino acids, said surface mated to a body having a cavity for sealing said cavity such that said linker molecules are in fluid communication with said cavity; wherein said cavity is being less than 1000 μm deep;

an inlet port and an outlet port, said inlet port and said outlet port being in communication with said cavity, wherein said cavity is less than 1000 μm deep;

a fluid flowing means coupled to said inlet port positioned to deliver fluid to the cavity and

a means for heating said cavity.

* * * * *

EXHIBIT 4

US006576424B2

(12) **United States Patent**
Fodor et al.

(10) **Patent No.:** **US 6,576,424 B2**
(45) **Date of Patent:** ***Jun. 10, 2003**

(54) **ARRAYS AND METHODS FOR DETECTING NUCLEIC ACIDS**

(75) Inventors: **Stephen P. A. Fodor**, Palo Alto, CA (US); **Dennis W. Solas**, San Francisco, CA (US); **William J. Dower**, Menlo Park, CA (US)

(73) Assignee: **Affymetrix Inc.**, Santa Clara, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 249 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **09/768,626**

(22) Filed: **Jan. 25, 2001**

(65) **Prior Publication Data**

US 2002/0164590 A1 Nov. 7, 2002

Related U.S. Application Data

(63) Continuation of application No. 09/670,563, filed on Sep. 27, 2000, which is a continuation of application No. 09/362,089, filed on Jul. 28, 1999, which is a division of application No. 09/056,927, filed on Apr. 8, 1998, now Pat. No. 6,197,506, which is a continuation of application No. 08/670,118, filed on Jun. 25, 1996, now Pat. No. 5,800,992, which is a division of application No. 08/168,904, filed on Dec. 15, 1993, now abandoned, which is a continuation of application No. 07/624,114, filed on Dec. 6, 1990, now abandoned.

(51) **Int. Cl.**⁷ **C12Q 1/68**; C12M 1/34; C07H 21/04; C07H 21/02

(52) **U.S. Cl.** **435/6**; 435/287.2; 435/288.3; 536/23.1; 536/24.3; 536/24.31; 536/24.32

(58) **Field of Search** 435/6, 287.2, 288.3; 536/23.1, 24.3, 24.31, 24.32

(56) **References Cited**

U.S. PATENT DOCUMENTS

3,849,137 A 11/1974 Barzynski et al.
3,862,056 A 1/1975 Hartman
3,939,350 A 2/1976 Kronick et al.
4,072,576 A 2/1978 Arwin et al.
4,180,739 A 12/1979 Abu-Shumays
4,238,757 A 12/1980 Schenck
4,269,933 A 5/1981 Pazos
4,314,821 A 2/1982 Rice
4,327,073 A 4/1982 Huang
4,339,528 A 7/1982 Goldman
4,342,905 A 8/1982 Fujii et al.
4,373,071 A 2/1983 Itakura
4,405,771 A 9/1983 Jagur
4,444,878 A 4/1984 Paulus
4,444,892 A 4/1984 Malmros
4,448,534 A 5/1984 Wertz et al.
4,458,066 A 7/1984 Caruthers et al.
4,483,920 A 11/1984 Gillespie et al.
4,500,707 A 2/1985 Caruthers et al.
4,516,833 A 5/1985 Fusek
4,517,338 A 5/1985 Urdea et al.
4,537,861 A 8/1985 Elings et al.

4,542,102 A 9/1985 Dattagupta et al.
4,555,490 A 11/1985 Merrill
4,562,157 A 12/1985 Lowe et al.
4,569,967 A 2/1986 Kornreich et al.
4,580,895 A 4/1986 Patel
4,584,277 A 4/1986 Ullman
4,613,566 A 9/1986 Potter
4,624,915 A 11/1986 Schindler et al.
4,626,684 A 12/1986 Landa
4,631,211 A 12/1986 Houghten
4,637,861 A 1/1987 Krull et al.
4,677,054 A 6/1987 White et al.
4,681,859 A 7/1987 Kramer
4,683,202 A 7/1987 Mullis
4,689,405 A 8/1987 Frank et al.
4,704,353 A 11/1987 Humphries et al.
4,711,955 A 12/1987 Ward et al.
4,713,326 A 12/1987 Dattagupta et al.
4,713,347 A 12/1987 Mitchell et al.
4,719,615 A 1/1988 Feyrer et al.
4,722,906 A 2/1988 Guire
4,728,502 A 3/1988 Hamill
4,728,591 A 3/1988 Clark et al.
4,731,325 A 3/1988 Palva et al.
4,755,458 A 7/1988 Rabbani et al.
4,762,881 A 8/1988 Kauer

(List continued on next page.)

FOREIGN PATENT DOCUMENTS

DE 2242394 3/1974
DE 3440141 5/1986
DE 3505287 3/1988

(List continued on next page.)

OTHER PUBLICATIONS

Sequencing by Hybridization Workshop, listing of participants and workshop presentation summaries (1991).
“A Sequencing Reality Check,” *Science*, 242:1245 (1988).
“Affymax raises \$25 million to develop high-speed drug discovery system,” *Biotechnology News*, 10(3):7-8 (1990).
“Preparation of fluorescent-labeled DNA and its use as a probe in molecular hybridization,” *Bioorg Khim*, 12(11):1508-1513 (1986).

Abbott et al., “Manipulation of the Wettability of Surfaces on the 0.1-to 1-Micrometer Scale Through Micromachining and Molecular Self-Assembly,” *Science*, 257:1380-1382 (1992).

(List continued on next page.)

Primary Examiner—Stephanie W. Zitomer
(74) *Attorney, Agent, or Firm*—Morgan, Lewis & Bockius LLP

(57) **ABSTRACT**

The present invention provides methods and apparatus for sequencing, fingerprinting and mapping biological macromolecules, typically biological polymers. The methods make use of a plurality of sequence specific recognition reagents which can also be used for classification of biological samples, and to characterize their sources.

64 Claims, 2 Drawing Sheets

US 6,576,424 B2

Page 2

U.S. PATENT DOCUMENTS					
4,777,019	A	10/1988	Dandekar	5,436,327	A 7/1995 Southern et al.
4,780,504	A	10/1988	Buendia et al.	5,445,934	A 8/1995 Fodor et al.
4,786,170	A	11/1988	Groeblner	5,447,841	A 9/1995 Gray et al.
4,786,684	A	11/1988	Glass	5,451,505	A 9/1995 Dollinger
4,794,150	A	12/1988	Steel	5,474,796	A 12/1995 Brennan
4,808,508	A	2/1989	Platzner	5,486,452	A 1/1996 Gordon et al.
4,810,869	A	3/1989	Yabe et al.	5,489,507	A 2/1996 Chehab
4,811,062	A	3/1989	Tabata et al.	5,489,678	A 2/1996 Fodor et al.
4,812,512	A	3/1989	Buendia et al.	5,492,806	A 2/1996 Drmanac et al.
4,820,630	A	4/1989	Taub	5,510,270	A 4/1996 Fodor et al.
4,822,566	A	4/1989	Newman	5,525,464	A 6/1996 Drmanac et al.
4,833,092	A	5/1989	Geysen	5,527,681	A 6/1996 Holmes
4,844,617	A	7/1989	Kelderman et al.	5,552,270	A 9/1996 Khrapko et al.
4,846,552	A	7/1989	Veldkamp et al.	5,556,961	A 9/1996 Foote et al.
4,849,513	A	7/1989	Smith et al.	5,561,071	A 10/1996 Hollenberg et al.
4,855,225	A	8/1989	Fung et al.	5,565,324	A 10/1996 Still
4,865,990	A	9/1989	Stead et al.	5,567,809	A 10/1996 Apple
4,868,103	A	9/1989	Stavrianopoulos et al.	5,571,639	A 11/1996 Hubbell et al.
4,874,500	A	10/1989	Madou et al.	5,573,905	A 11/1996 Lerner
4,886,741	A	12/1989	Schwartz	5,593,839	A 1/1997 Hubbell et al.
4,888,278	A	12/1989	Singer et al.	5,604,097	A 2/1997 Brenner
4,923,901	A	5/1990	Koester et al.	5,635,400	A 6/1997 Brenner
4,925,785	A	5/1990	Wang et al.	5,641,634	A 6/1997 Mandeck
4,946,942	A	8/1990	Fuller et al.	5,653,939	A 8/1997 Hollis et al.
4,965,188	A	10/1990	Mullis	5,654,413	A 8/1997 Brenner
4,973,493	A	11/1990	Guire	5,667,667	A 9/1997 Southern
4,979,959	A	12/1990	Guire	5,667,972	A 9/1997 Drmanac et al.
4,981,783	A	1/1991	Augenlicht	5,690,894	A 11/1997 Pinkel
4,981,985	A	1/1991	Kaplan et al.	5,695,940	A 12/1997 Drmanac et al.
4,984,100	A	1/1991	Takayama et al.	5,698,393	A 12/1997 Maciszek et al.
4,987,065	A	1/1991	Stavrianopoulos et al.	5,700,637	A 12/1997 Southern
4,988,617	A	1/1991	Landegren et al.	5,707,806	A 1/1998 Shuber
4,992,383	A	2/1991	Farnsworth	5,744,305	A * 4/1998 Fodor et al. 435/6
4,994,373	A	2/1991	Stavrianopoulos et al.	5,751,629	A 5/1998 Nova
5,002,867	A	3/1991	Macevitz	5,770,367	A 6/1998 Southern
5,021,550	A	6/1991	Zeiger	5,777,888	A 7/1998 Rine et al.
5,026,773	A	6/1991	Steel	5,800,992	A * 9/1998 Fodor et al. 435/6
5,026,840	A	6/1991	Dattagupta et al.	5,804,563	A 9/1998 Still
5,028,525	A	7/1991	Gray et al.	5,807,522	A 9/1998 Brown et al.
5,043,265	A	8/1991	Tanke et al.	5,807,683	A 9/1998 Brenner
5,047,524	A	9/1991	Andrus et al.	5,830,645	A 11/1998 Pinkel et al.
5,075,077	A	12/1991	Durley	5,843,767	A 12/1998 Beattie
5,077,210	A	12/1991	Eigler	5,846,708	A 12/1998 Hollis et al.
5,079,600	A	1/1992	Schnur et al.	5,846,719	A 12/1998 Brenner
5,081,584	A	1/1992	Omichinski et al.	5,863,722	A 1/1999 Brenner
5,082,830	A	1/1992	Brakel et al.	5,871,697	A 2/1999 Rothberg et al.
5,091,652	A	2/1992	Mathies et al.	5,871,928	A * 2/1999 Fodor et al. 435/6
5,112,962	A	5/1992	Letsinger et al.	6,023,540	A 2/2000 Walt
5,141,813	A	8/1992	Nelson	6,054,270	A 4/2000 Southern
5,143,854	A	9/1992	Pirrung et al.	6,060,240	A 5/2000 Kamb
5,153,319	A	10/1992	Caruthers et al.	6,197,506	B1 * 3/2001 Fodor et al. 435/6
5,192,980	A	3/1993	Dixon et al.	FOREIGN PATENT DOCUMENTS	
5,200,051	A	4/1993	Cozzette et al.	DE	3722958 1/1989
5,202,231	A	4/1993	Drmanac et al.	EP	046 083 2/1982
5,206,137	A	4/1993	Ip et al.	EP	088 636 9/1983
5,215,882	A	6/1993	Bahl et al.	EP	103 197 3/1984
5,215,889	A	6/1993	Schultz	EP	127 438 12/1984
5,232,829	A	8/1993	Longiaru et al.	EP	063 810 3/1986
5,235,028	A	8/1993	Barany et al.	EP	194 132 9/1986
5,242,974	A	9/1993	Holmes	EP	228 075 7/1987
5,252,743	A	10/1993	Barrett et al.	EP	245 662 11/1987
5,256,549	A	10/1993	Urdea et al.	EP	268 237 5/1988
5,258,506	A	11/1993	Urdea et al.	EP	281 927 9/1988
5,306,641	A	4/1994	Saccocio	EP	228 310 10/1988
5,310,893	A	5/1994	Erllich et al.	EP	288 310 10/1988
5,324,633	A	6/1994	Fodor et al.	EP	304 202 2/1989
5,348,855	A	9/1994	Dattagupta et al.	EP	307 476 3/1989
5,384,261	A	1/1995	Winkler et al.	EP	319 012 6/1989
5,405,783	A	4/1995	Pirrung et al.	EP	328 256 8/1989
5,424,186	A	6/1995	Fodor et al.	EP	333 561 9/1989

US 6,576,424 B2

Page 3

EP	337 498	10/1989
EP	386 229	4/1990
EP	373 203	6/1990
EP	392 546	10/1990
EP	173 339	1/1992
EP	171 150	3/1992
EP	237 362	3/1992
EP	185 547	6/1992
EP	260 634	6/1992
EP	232 967	4/1993
EP	235 726	5/1993
EP	476 014	8/1994
EP	225 807	10/1994
EP	717 113	6/1996
EP	0 721 016	7/1996
EP	848 067	6/1998
EP	619 321	1/1999
FR	2559783	3/1988
GB	2 129 551	5/1984
GB	2156074	3/1988
GB	2196476	4/1988
GB	8810400	5/1988
GB	2233654	1/1991
GB	2248840	9/1992
JP	49-110601	10/1974
JP	60-248669	12/1985
JP	63-084499	4/1988
JP	63-223557	9/1988
JP	1-233447	9/1989
NO	913186	8/1991
WO	WO 84/03151	8/1984
WO	WO 84/03564	9/1984
WO	WO 85/01051	3/1985
WO	WO 86/00991	2/1986
WO	WO 86/06487	11/1986
WO	WO 88/04777	6/1988
WO	WO 89/05616	6/1989
WO	WO 89/08834	9/1989
WO	WO 89/10977	11/1989
WO	WO 89/11548	11/1989
WO	WO 89/12819	12/1989
WO	WO 90/00626	1/1990
WO	WO 90/00887	2/1990
WO	WO 90/15070	2/1990
WO	WO 90/03382	4/1990
WO	WO 90/04652	5/1990
WO	WO 91/04266	4/1991
WO	WO 91/07087	5/1991
WO	WO 92/16655	1/1992
WO	WO 92/10092	6/1992
WO	WO 92/10588	6/1992
WO	WO 93/02992	2/1993
WO	WO 93/09668	5/1993
WO	WO 88/01302	6/1993
WO	WO 93/11262	6/1993
WO	WO 93/17126	9/1993
WO	WO 93/22456	11/1993
WO	WO 93/22480	11/1993
WO	WO 95/00530	1/1995
WO	WO 95/11995	5/1995
WO	WO 95/33846	12/1995
WO	WO 96/23078	8/1996
WO	WO 97/10365	3/1997
WO	WO 97/17317	5/1997
WO	WO 97/19410	5/1997
WO	WO 97/27317	7/1997
WO	WO 97/29212	8/1997
WO	WO 98/31836	7/1998
WO	WO 99/60007	11/1999

OTHER PUBLICATIONS

Adams et al., "Complementary DNA Sequencing: Expressed Sequence Tags and Human Genome Project," *Science*, 252(5013):1651-1656 (1991).

Adams et al., "Photolabile Chelators That "Cage" Calcium with Improved Speed of Release and Pre-Photolysis Affinity," *J. Gen. Physiol.*, p. 9a (12/86).

Adams et al., "Biologically Useful Chelators That Take Up Ca²⁺ upon Illumination," *J. Am. Chem. Soc.*, 111:7957-7968 (1989).

Amit et al., "Photosensitive Protecting Groups of Amino Sugars and Their Use in Glycoside Synthesis. 2-Nitrobenzylloxycarbonylamino and 6-Nitroveratryloxycarbonylamino Derivatives," *J. Org. Chem.*, 39(2):192-196 (1974).

Amit et al., "Photosensitive Protecting Groups—A Review," *Israel J. Chem.*, 12(1-2):103-113 (1974).

Applied Biosystems, Model 431A Peptide Synthesizer User's manual, Sections 2 and 6, (8/15/89).

Ajayaghosh et al., "Solid-Phase Synthesis of N-Methyl- and N-Ethylamides of Peptides Using Photolytically Detachable ((3-Nitro-4((alkylamino)methyl)benzamido)methyl)polystyrene Resin," *J. Org. Chem.*, 55(9):2826-2829 (1990).

Ajayaghosh et al., "Solid-phase synthesis of C-terminal peptide amides using a photoremovable α -methylphenacylamido anchoring linkage," *Proc. Ind. Natl. Sci (Chem.Sci.)*, 100(5):389-396 (1988).

Ajayaghosh et al., "Polymer-supported Solid-phase Synthesis of C-Terminal Peptide N-Methylamides Using a Modified Photoremovable 3-Nitro-4-N-methylaminomethylpolystyrene Support," *Ind. J. Chem.*, 27B:1004-1008 (1988).

Ajayaghosh et al., "Polymer-Supported Synthesis of Protected Peptide Segments on a Photosensitive o-Nitro(α -Methyl)Bromobenzyl Resin," *Tetrahedron*, 44(21):6661-6666 (1988).

Arnold et al., "A Novel Universal Support for DNA & RNA Synthesis," abstract from *Federation Proceedings*, 43(7): abstract No. 3669 (1984).

Atherton et al., *Solid Phase Peptide Synthesis: A Practical Approach*, IRL Press, (1989), tbl. of cont., pp. vii-ix.

Augenlicht et al., "Cloning and Screening of Sequences Expressed in a Mouse Colon Tumor," *Cancer Research*, 42:1088-1093 (1982).

Augenlicht et al., "Expression of Cloned Sequences in Biopsies of Human Colonic Tissue and in Colonic Carcinoma Cells Induced to Differentiate in Vitro," *Cancer Res.*, 47:6017-6021 (1987).

Bains, W., "Hybridization Methods for DNA Sequencing," *Genomics*, 11(2):294-301 (1991).

Bains et al., "A Novel Method for Nucleic Acid Sequence Determination," *J. Theor. Biol.*, 135:303-307 (1988).

Bains, W., "Alternative Routes Through the Genome," *Biotechnology*, 8:1251-1256 (1988).

Balachander et al., "Functionalized Siloxy-Anchored Monolayers with Exposed Amino, Azido, Bromo, or Cyano Groups," *Tetrahed. Lett.*, 29(44):5593-5594 (1988).

Baldwin et al., "New Photolabile Phosphate Protecting Groups," *Tetrahed.*, 46(19):6879-6884 (1990).

Barltrop et al., "Photosensitive Protective Groups," *Chemical Communications*, pp. 822-823 (1966).

Barinaga, M., "Will 'DNA Chip' Speed Genome Initiative," *Science*, 253:1489 (1985).

US 6,576,424 B2

Page 4

- Bart et al., "Microfabricated Electrohydrodynamic Pumps," *Sensors and Actuators*, A21-A23:193-197 (1990).
- Bartsh et al., "Cloning of mRNA sequences from the human colon: Preliminary characterisation of defined mRNAs in normal and neoplastic tissues," *Br.J.Can.*, 54:791-798 (1986).
- Baum, R., "Fledgling firm targets drug discovery process," *Chem. Eng. News*, p. 10-11 (1990).
- Beltz et al., "Isolation of Multigene Families and Determination of Homologies by Filter Hybridization Methods," *Methods in Enzymology*, 100:266-285 (1983).
- Benschop, Chem. Abstracts 114(26):256643 (1991).
- Bhatia et al., "New Approach To Producing Patterned Biomolecular Assemblies," *J. American Chemical Society*, 114:4432-4433 (1992).
- Biorad Chromatography Electrophoresis Immunochemistry Molecular Biology HPLC catalog M 1987 pp. 182.
- Blawas et al., "Step-and-Repeat Photopatterning of Protein Features Using Caged-Biotin-BSA: Characterization and Resolution," *Langmuir*, 14(15):4243-4250 (1998).
- Blawas, A.S., "Photopatterning of Protein Features using Caged-biotin-Bovine Serum Albumin," dissertation for Ph.D at Duke University in 1998.
- Bos et al., "Amino-acid substitutions at codon 13 of the N-ras oncogene in human myeloid leukaemia," *Nature*, 315:726-730 (1985).
- Boyle et al., "Differential distribution of long and short interspersed element sequences in the mouse genome: Chromosome karyotyping by fluorescence in situ hybridization," *PNAS*, 87:7757-7761 (1990).
- Brock et al., "Rapid fluorescence detection of in situ hybridization with biotinylated bovine herpesvirus-1 DNA probes," *J.Veterinary Diagnostic Invest.*, 1:34-38 (1989).
- Burgi et al., "Optimization in Sample Stacking for High-Performance Capillary Electrophoresis," *Anal. Chem.*, 63:2042-2047 (1991).
- Cameron et al., "Photogeneration of Organic Bases from o-Nitrobenzyl-Derived Carbamates," *J. Am. Chem. Soc.*, 113:4303-4313 (1991).
- Carrano et al., "A High-Resolution, Fluorescence-Based, Semiautomated Method for DNA Fingerprinting," *Genomics*, 4:129-136 (1989).
- Caruthers, M.H., "Gene Synthesis Machines: DNA Chemistry and Its Uses," *Science*, 230:281-285 (1985).
- Chatterjee et al., "Inducible Alkylation of DNA Using an Oligonucleotide-Quinone Conjugate," *Am. J. Chem. Soc.*, 112:6397-6399 (1990).
- Chee et al., "Accessing Genetic Information with High-Density DNA Arrays," *Science*, 274:610-614 (1996).
- Chehab et al., "Detection of sickle cell anaemia mutation by colour DNA amplification," *Lancet*, 335:15-17 (1990).
- Chehab et al., "Detection of specific DNA sequences by fluorescence amplification: A color complementation assay," *PNAS*, 86:9178-9182 (1989).
- Clevite Corp., Piezoelectric Technology, Data for Engineers.
- Corbett et al., "Reaction of Nitroso Aromatics with Glyoxylic Acid. A New Path to Hydroxamic Acids," *J. Org. Chem.*, 45:2834-2839 (1980).
- Craig et al., "Ordering of cosmid clones covering the Herpes simplex virus type 1 (HSV-1) genome: a test case for fingerprinting by hybridization," *Nuc. Acid. Res.*, 18(9):2653-2660 (1990).
- Cummings et al., "Photoactivable Fluorophores. 1. Synthesis and Photoactivation of o-Nitrobenzyl-Quenched Fluorescent Carbamates," *Tetrahedron Letters*, 29(1):65-68 (1988).
- Diggelmann, "Investigating the VLSIPS synthesis process," 9/9/94.
- Di Mauro et al., "DNA Technology in Chip Construction," *Adv. Mater.*, 5(5):384-386 (1993).
- Drmanac et al., "Partial Sequencing by Oligo-Hybridization Concept and Applications in Genome Analysis," 1st Int. Conf. Electrophor., Supercomp., Hum. Genome pp. 60-74 (1990).
- Drmanac et al., "Sequencing by Oligonucleotide Hybridization: A Promising Framework in Decoding of the Genome Program?," 1st Int. Conf. Electrophor., Supercomp., Hum. Genome pp. 47-59 (1990).
- Drmanac et al., "Laboratory Methods, Reliable Hybridization of Oligonucleotides as Short as Six Nucleotides," *DNA and Cell Biol.*, 9(7):527-534 (1990).
- Drmanac et al., "Sequencing of Megabase Plus DNA by Hybridization: theory of the Method," *Genomics*, 4:114-128 (1989).
- Dramanac et al., "Sequencing of Megabase Plus DNA by Hybridization: Theory of the Method," abstract of presentation given at Cold Spring Harbor Symposium on Genome Mapping and Sequencing, 4/27/88 thru 5/1/88.
- Dulcey et al., "Deep UV Photochemistry of Chemisorbed Monolayers: Patterned Coplanar Molecular Assemblies," *Science*, 252:551-554 (1991).
- Duncan et al., "Affinity Chromatography of a Sequence-Specific DNA Binding Protein Using Teflon-Linked Oligonucleotides," *Analytical Biochemistry*, 169:104-108 (1988).
- Effenhauser et al., "Glass Chips for High-speed Capillary Electrophoresis Separations with Submicrometer Plate Heights," *Anal. Chem.*, 65:2637-2642 (1993).
- Effenhauser et al., "High-Speed Separation of Antisense Oligonucleotides on a Micromachined Capillary Electrophoresis Device," *Anal. Chem.*, 66:2949-2953 (1994).
- Ekins et al., "High Specific Activity Chemiluminescent and Fluorescent Markers: their Potential Application to High Sensitivity and 'Multi-analyte' Immunoassays," *J. Bioluminescence Chemiluminescence*, 4:59-78 (1989).
- Ekins et al., "Development of Microspot Multi-Analyte Ratiometric Immunoassay Using dual Fluorescent-Labelled Antibodies," *Anal. Chemica Acta*, 227:73-96 (1989).
- Ekins et al., "Multianalyte Microspot Immunoassay-Microanalytical 'Compact Disk' of the Future," *Clin. Chem.*, 37(11):1955-1967 (1991).
- Ekins, R.P., "Multi-Analyte immunoassay*," *J. Pharmaceutical Biomedical Analysis*, 7(2):155-168 (1989).
- Ekins et al., "Fluorescence Spectroscopy and its Application to a New Generation of High Sensitivity, Multi-Microspot, Multianalyte, Immunoassay," *Clin. Chim. Acta*, 194:91-114 (1990).
- Evans et al., "Microfabrication for Automation of Molecular processes in Human Genome Analysis," *Clin. Chem.*, 41(11):1681 (1995).
- Evans et al., "Physical mapping of complex genomes by cosmid multiplex analysis," *PNAS*, 86:5030-5034 (1989).
- Ezaki et al., "Small-Scale DNA Preparation for Rapid Genetic Identification of Campylobacter Species without Radioisotope," *Microbiol. Immunology*, 32(2):141-150 (1988).

US 6,576,424 B2

Page 5

- Fan et al., "Mapping small DNA sequences by fluorescence in situ hybridization directly on banded metaphase chromosomes," *PNAS*, 87(16):6223-6227 (1990).
- Fan et al., "Micromachining of Capillary Electrophoresis Injectors and Separators on Glass Chips and Evaluation of Flow at Capillary Intersections," *Anal. Chem.*, 66:177-184 (1994).
- Fettingner et al., "Stacked modules for micro flow systems in chemical analysis: concept and studies using an enlarged model," *Sensors and Actuators*, B17:19-25 (1993).
- Flanders et al., "A new interferometric alignment technique," *App. Phys. Ltrs.*, 31(7):426-429 (1977).
- Fodor et al., "Multiplexed biochemical assays with biological chips," *Nature*, 364:555-556 (1993).
- Fodor et al., "Light-directed, Spatially Addressable Parallel Chemical Synthesis," *Science*, 251:767-773 (1991).
- Forman et al., "Thermodynamics of Duplex Formation and Mismatch Discrimination on Photolithographically Synthesized Oligonucleotide Arrays," chapter 13 pp. 206-228 from *Molecular Modeling of Nucleic Acids*, ACS Symposium Series 682, 4/13-17/97, Leontis et al., eds.
- Frank et al., "Simultaneous Multiple Peptide Synthesis Under Continuous flow Conditions on Cellulose Paper Discs as Segmental Solid Supports," *Tetrahedron*, 44(19):6031-6040 (1988).
- Frank et al., "Automation of DNA Sequencing Reactions and Related Techniques: A Workstation for Micromanipulation of Liquids," *Bio/Technology*, 6:1211-1212 (1988).
- Frank et al., "Simultaneous Synthesis and Biological Applications of DNA Fragments: An Efficient and Complete Methodology," *Methods in Enzymology*, 154:221-250 (1987).
- Fuhr et al., "Travelling wave-driven microfabricated electrohydrodynamic pumps for liquids," *J. Micromech. Microeng.*, 4:217-226 (1994).
- Fuller et al., "Urethane-Protected Amino Acid N-Carboxy Anhydrides and Their Use in Peptide Synthesis," *J. Amer. Chem. Soc.*, 112(20):7414-7416 (1990).
- Furka et al., "General method for rapid synthesis of multi-component peptide mixtures," *Int. J. Peptide Protein Res.*, 37:487-493 (1991).
- Furka et al., "Cornucopia of Peptides by Synthesis," 14th Int. Congress of Biochem. abst.# FR:013, 7/10-15/88 Prague, Czechoslovakia.
- Furka et al., "More Peptides by Less Labour," abst. 288, Int. Symp. Med. Chem., Budapest Hungary 8/15-19/88.
- Gait, eds., pp. 1-115 from *Oligonucleotide Synthesis: A Practical Approach*, IRL Press, (1984).
- Gazard et al., "Lithographic Technique Using Radiation-Induced Grafting of Acrylic Acid into Poly(Methyl Methacrylate) Films," *Polymer Engineering and Science*, 20(16):1069-1072 (1980).
- Gergen et al., "Filter replicas and permanent collections of recombinant DNA plasmids," *Nuc. Acids Res.*, 7(8):2115-2137 (1979).
- Getzoff et al., "Mechanisms of Antibody Binding to a Protein," *Science*, 235:1191-1196 (1987).
- Geysen et al., "Strategies for epitope analysis using peptide synthesis," *J. Immunol. Meth.*, 102:259-274 (1987).
- Geysen et al., "Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid," *PNAS*, 81:3998-4002 (1984).
- Geysen et al., "A synthetic strategy for epitope mapping," from *Peptides: Chem. & Biol.*, Proc. of 10th Am. Peptide Symp., 5/23-28/87, pp. 519-523, (1987).
- Geysen, "Antigen-antibody interactions at the molecular level: adventures in peptide synthesis," *Immunol. Today*, 6(12):364-369 (1985).
- Geysen et al., "Cognitive Features of Continuous Antigenic Determinants," from *Synthetic Peptides: Approaches to Biological Probes*, pp. 19-30, (1989).
- Geysen et al., "Chemistry of Antibody Binding to a Protein," *Science*, 235:1184-1190 (1987).
- Geysen et al., "The delineation of peptides able to mimic assembled epitopes," 1986 CIBA Symp., pp. 130-149.
- Geysen et al., "Cognitive Features of Continuous Antigenic Determinants," *Mol. Recognit.*, 1(1):1-10 (1988).
- Geysen et al., "A Prio Ri Delineation of a Peptide Which Mimics A Discontinuous Antigenic Determinant," *Mol. Immunol.*, 23(7):709-715 (1986).
- Gilon et al., "Backbone Cyclization: A New Method for Conferring Conformational Constraint on Peptides," *Biopolymers*, 31(6):745-750 (1991).
- Gingeras et al., "Hybridization properties of immobilized nucleic acids," *Nuc. Acids Res.*, 15(13):5373-5390 (87).
- Gummerlock et al., "RAS Enzyme-Linked Immunoblot Assay Discriminates p21 Species: A Technique to Dissect Gene Family Expression," *Anal. Biochem.*, 180:158-168 (1989).
- Gurney et al., "Activation of a potassium current by rapid photochemically generated step increases of intracellular calcium in rat sympathetic neurons," *PNAS*, 84:3496-3500 (1987).
- Haase et al., "Detection of Two Viral Genomes in Single Cells by Double-Label Hybridization in Situ and Color Microradioautography," *Science*, 227:189-192 (1985).
- Hacia, et al., "Two color hybridization analysis using high density oligonucleotide arrays and energy transfer dyes," *Nuc. Acids Res.*, 26(16):3865-3866 (1998).
- Hack, M.L., "Conics Formed to Make Fluid & Industrial Gas Micromachines," *Genetic Engineering News*, 15(18):1, 29 (1995).
- Hagedorn et al., "Pumping of Water Solutions in Microfabricated Electrohydrodynamic Systems," from Micro Electro Mechanical Systems conference in Travemunde Germany (1992).
- Hames et al., *Nuclear acid hybridization, a practical approach*, cover page and table of contents (1985).
- Hanahan et al., "Plasmid Screening at High Colony Density," *Meth. Enzymology*, 100:333-342 (1983).
- Hanahan et al., "Plasmid screening at high colony density," *Gene*, 10:63-67 (1980).
- Haridasan et al., "Peptide Synthesis using Photolytically Cleavable 2-Nitrobenzylloxycarbonyl Protecting Group," *Proc. Indian Natn. Sci. Acad.*, 53A(6):717-728 (1987).
- Harrison et al., "Capillary Electrophoresis and Sample Injection Systems Integrated on a Planar Glass Chip," *Anal. Chem.*, 64:1926-1932 (1992).
- Harrison et al., "Micromachining a Minaturized Capillary Electrophoresis-Based Chemical Analysis System on a Chip," *Science*, 261:895-897 (1993).
- Harrison et al., "Towards minaturized electrophoresis and chemical analysis systems on silicon: an alternative to chemical sensors*," *Sensors and Actuators*, B10:107-116 (1993).

US 6,576,424 B2

Page 6

- Harrison et al., "Rapid Separation of Fluorescein Derivatives using a micromachined capillary electrophoresis system," *Analytica Chimica Acta*, 283:361-366 (1993).
- Hellberg et al., "Minimum analogue peptide sets (MAPS) for quantitative structure-activity relationships," *Int. J. Peptide Protein Res.*, 37:414-424 (1991).
- Hilser et al., "Protein and peptide mobility in capillary zone electrophoresis, A comparison of existing models and further analysis," *J. Chromatography*, 630:329-336 (1993).
- Ho et al., "Highly Stable Biosensor Using an Artificial Enzyme," *Anal.Chem.*, 59:536-537 (1987).
- Hochgeschwender et al., "Preferential expression of a defined T-cell receptor β -chain gene in hapten-specific cytotoxic T-cell clones," *Nature*, 322:376-378 (1986).
- Hodgson, J., "Assays A La Photolithography," *Biotech.*, 9:419 (1991).
- Hopman et al., "Bi-color detection of two target DNAs by non-radioactive in situ hybridization*," *Histochem.*, 85:1-4 (1986).
- Iwamura et al., "1-Pyrenylmethyl Esters, Photolabile Protecting Groups for Carboxylic Acids," *Tetrahedon Ltrs.*, 28(6):679-682 (1987).
- Iwamura et al., "1-(α -Diazobenzyl)pyrene: A Reagent for Photolabile and Fluorescent Protection of Carboxyl Groups of Amino Acids and Peptides," *Synlett*, p. 35-36 (1991).
- Jacobson et al., "Effects of Injection Schemes and Column Geometry on the Performance of Microchip Electrophoresis Devices," *Anal. Chem.*, 66:1107-1113 (1994).
- Jacobsen et al., "Open Channel Electrochromatography on a Microchip," *Anal. chem.*, 66:2369-2373 (1994).
- Jacobson et al., "Microchip Capillary Electrophoresis with an Integrated Postcolumn Reactor" *Anal. Chem.*, 66:3472-3476 (1994).
- Jacobson et al., "Precolumn Reactions with Electrophoretic Analysis Integrated on a Microchip," *Anal. Chem.*, 66:4127-4132 (1994).
- Jacobson et al., "Microfabricated chemical measurement systems," *Nature Medicine*, 1(10):1093-1096 (1995).
- Jacobsen et al., "Fused Quartz Substrates for Microchip Electrophoresis," *Anal. chem.*, 67:2059-2063 (1995).
- Jacobson et al., "High-Speed Separations on a Microchip," *Anal. Chem.*, 66:1114-1118 (1994).
- Jacobson et al., "Microchip electrophoresis with sample stacking," *Electrophoresis*, 16:481-486 (1995).
- Jayakumari, "Peptide synthesis in a triphasic medium catalyzed by papain immobilized on a crosslinked polystyrene support," *Indian J. Chemistry*, 29B:514-517 (1990).
- Kaiser et al., "Peptide and Protein Synthesis by Segment Synthesis-Condensation," *Science*, 243:187-192 (1989).
- Kaplan et al., "Photolabile chelators for the rapid photorelease of divalent cations," *PNAS*, 85:6571-6575 (1988).
- Karube, "Micro-biosensors based on silicon fabrication technology," chapter 25 from *Biosensors: Fundamentals and Applications*, Turner et al., eds., Oxford Publ., 1987, pp. 471-480 (1987).
- Kates et al., "A Novel, Convenient, Three-dimensional Orthogonal Strategy for Solid-Phase Synthesis of Cyclic Peptides 1-3," *Tetrahed. Letters*, 34(10):1549-1552 (1993).
- Kerkof et al., "A Procedure for Making Simultaneous Determinations of the Relative Levels of Gene Transcripts in Tissues or Cells," *Anal. Biochem.*, 188:349-355 (1990).
- Khrapko et al., "An Oligonucleotide hybridization approach to DNA sequencing," *FEBS Lett.*, 256(1,2):118-122 (1989).
- Kievits et al., "Rapid subchromosomal localization of cosmid by nonradioactive in situ hybridization," *Cytogenetics Cell Genetics*, 53(2-3):134-136 (1990).
- Kimura et al., "An Immobilized Enzyme Membrane Fabrication Method using an Ink Jet Nozzle," *Biosensors*, 4:41-52 (1988).
- Kimura et al., "An Integrated SOS/FET Multi-Biosensor," *Sensors & Actuators*, 9:373-387 (1986).
- Kitazawa et al., "In situ DNA-RNA hybridization using in vivo bromodeoxyuridine-labeled DNA probe," *Histochemistry*, 92:195-199 (1989).
- Kleinfeld et al., "Controlled Outgrowth of Dissociated Neurons on Patterned Substrates," *J. Neurosci.*, 8(11):4098-4120 (1988).
- Knight, P., "Materials and Methods/Microsequencers for Proteins and Oligosaccharides," *Bio/Techn.*, 7:1075-76 (1989).
- Kohara et al., "The Physical Map of the Whole *E. coli* Chromosome: Application of a New Strategy for Rapid Analysis and Sorting of a Large Genomic Library," *Cell*, 50:495-508 (1987).
- Krile et al., "Multiplex holography with chirp-modulated binary phase-coded reference-beam masks," *Applied Opt.*, 18(1):52-56 (1979).
- Labat, I., "Subfragments as an informative characteristic of the DNA molecule—computer simulation," research report submitted to the University of Belgrade College of Natural Sciences and Mathematics, (1988).
- Lainer et al., "Human Lymphocyte Subpopulations Identified by Using Three-Color Immunofluorescence and Flow Cytometry Analysis: Correlation of Leu-2, Leu-3, Leu-7, Leu-8, and Leu-11 Clonally Surface Antigen Expression," *Journal of Immunology*, 132(1):151-156 (1984).
- Lam et al., "A new type of synthetic peptide library for identifying ligand-binding activity," *Nature*, 354:82-84 (1991).
- Laskey et al., "Messenger RNA prevalence in sea urchin embryos measured with cloned cDNAs," *PNAS*, 77(9):5317-5321 (1980).
- Lee et al., "synthesis of a Polymer Surface Containing Covalently Attached Triethoxysilane Functionality: Adhesion to Glass," *Macromolecules*, 21:3353-3356 (1988).
- Lehrach et al., "Labelling oligonucleotides to high specific activity (1)," *Nuc. Acids Res.*, 17(12):4605-4610 (89).
- Lehrach et al., "Phage Vectors—EMBL Series," *Meth. Enzymology*, 153:103-115 (1987).
- Levy, M.F., "Preparing Additive Printed Circuits," *IBM Tech. Discl. Bull.*, 9(11):1473 (1967).
- Lichter et al., "High-Resolution Mapping of Human Chromosome 11 by in Situ hybridization with Cosmid Clones," *Science*, 247:64-69 (1990).
- Lichter et al., "Fluorescence in situ hybridization with Alu and L1 polymerase chain reaction probes for rapid characterization of human chromosomes in hybrid cell lines," *PNAS*, 87:6634-6638 (1990).
- Lichter et al., "Rapid detection of human chromosome 21 aberrations by in situ hybridization," *PNAS*, 85:9664-9668 (1988).
- Lichter et al., "Is non-isotopic in situ hybridization finally coming of age," *Nature*, 345:93-94 (1990).
- Lieberman et al., "A Light source Smaller Than the Optical Wavelength," *Science*, 247:59-61 (1990).
- Lipshutz et al., "Using Oligonucleotide Probe Arrays To Access Genetic Diversity," *BioTech.*, 19(3):442-7 (1995).

US 6,576,424 B2

Page 7

- Liu et al., "Sequential Injection Analysis in Capillary Format with an Electroosmotic Pump," *Talanta*, 41(11):1903-1910 (1994).
- Lockhart et al., "Expression monitoring by hybridization to high-density oligonucleotide arrays," *Nat. Biotech.*, 14:1675-1680 (1996).
- Logue et al., "General Approaches to Mask Design for Binary Optics," SPIE, 1052:19-24 (1989).
- Loken et al., "three-color Immunofluorescence Analysis of Leu Antigens on Human Peripheral Blood Using Two Lasers on a Fluorescence-Activated Cell Sorter," *Cytoetry*, 5:151-158 (1984).
- Love et al., "Screening of λ Library for Differentially Expressed Genes Using in Vitro Transcripts," *Anal. Biochem.*, 150:429-441 (1985).
- Lowe, C.R., "Biosensors," *Trends in Biotech.*, 2:59-65 (1984).
- Lowe, C.R., "An Introduction to the Concepts and Technology of Biosensors," *Biosensors*, 1:3-16 (1985).
- Lowe, C. R., Biotechnology and Crop Improvement and Protection, BCPC Publications, pp. 131-138 (1986).
- Lowe et al., "Solid-Phase Optoelectronic Biosensors," *Methods in Enzymology*, 137:338-347 (1988).
- Lowe, C.R., "Biosensors," *Phil. Tran. R. Soc. Lond.*, 324:487-496 (1989).
- Lu et al., "Differential screening of murine ascites cDNA libraries by means of in vitro transcripts of cell-cycle-phase-specific cDNA and digital image processing," *Gene*, 86:185-192 (1990).
- Lysov et al., "A new method for determining the DNA nucleotide sequence by hybridization with oligonucleotides," *Doklady Biochem.*, 303(1-6):436-438 (1989).
- Lysov et al., "DNA Sequencing by Oligonucleotide Hybridization," First International Conference on Electrophoresis, Supercomputing and the Human Genome, 4/10-13/90 p. 157.
- MacDonald et al., "A Rapid ELISA for Measuring Insulin in a Large Number of Research Samples," *Metabolism*, 38(5):450-452 (1989).
- Mairanovsky, V.G., "Electro-Deprotection—Electrochemical Removal of Protecting Groups**," *Angew. Chem. Int. Ed. Engl.*, 15(5):281-292 (1976).
- Manz et al., "Miniaturized Total Chemical Analysis Systems: a Novel Concept for Chemical Sensing," *Sensors and Actuators*, B1:244-248 (1990).
- Manz et al., "Micromachining of monocrystalline silicon and glass for chemical analysis systems, A look into next century's technology or just a fashionable craze?," *Trends in Analytical Chem.*, 10(5):144-149 (1991).
- Manz et al., "Planar chips technology for minaturization and integration of separation techniques into monitoring systems, Capillary electrophoresis on a chip," *J. Chromatography*, 593:253-258 (1992).
- Manz et al., "Planar Chips Technology for Miniaturization of Separation Systems: A Developing Perspective in Chemical Monitoring," chapter 1, 1-64 (1993).
- Manz et al., "Electroosmotic pumping and electrophoretic separations for miniaturized chemical analysis systems," *J. Micromech. Microeng.*, 4:257-265 (1994).
- Masiakowski et al., "Cloning of cDNA sequences of hormone-regulated genes from the MCF-7 human breast cancer cell line," *Nuc. Acids Res.*, 10(24):7895-7903 (1982).
- Matsumoto et al., "Preliminary Investigation of Micropumping Based on Electrical Control of Interfacial Tension," *IEEE*, pp. 105-110 (1990).
- Matsuzawa et al., "Containment and growth of neuroblastoma cells on chemically patterned substrates," *J. Neurosci. Meth.*, 50:253-260 (1993).
- McCray et al., "Properties and Uses of Photoreactive Caged Compounds," *Ann. Rev. Biophys. Biophys. Chem.*, 18:239-270 (1989).
- McGall et al., "The Efficiency of Light-Directed Synthesis of DNA Arrays on Glass Substrates," *J. American Chem. Soc.*, 119(2):5081-5090 (1997).
- McGillis, VLSI Technology, Sze, eds., Chapter 7, "Lithography," pp. 267-301 (1983).
- McMurray, J.S., "Solid Phase Synthesis of a Cyclic Peptide Using Fmoc Chemistry," *Tetrahedron Letters*, 32(52):7679-7682 (1991).
- Meinkoth et al., "Review: Hybridization of Nucleic Acids Immobilized on solid Supports," *Analytical Biochem.*, 138:267-284 (1984).
- Melcher et al., "Traveling-Wave Bulk Electroconvection Induced across a Temperature Gradient," *Physics of Fluids*, 10(6):1178-1185 (1967).
- Merrifield, R.B., "Solid Phase peptide Synthesis. 1. The Synthesis of a Tetrapeptide," *J.Am.Chem.Soc.*, 85:2149-2154 (1963).
- Michiels et al., "Molecular approaches to genome analysis: a strategy for the construction of ordered overlapping clone libraries," *CABIOS*, 3(3):203-10 (1987).
- Mirzabekov, A.D., "DNA sequencing by hybridization—a megasequencing method and a diagnostic tool?," *TIBTECH*, 12:27-32 (1994).
- Monaco et al., "Human Genome Linking with Cosmids and Yeast Artificial Chromosomes", abstract from CSHS, p. 50, (1989).
- Morita et al., "Direct pattern fabrication on silicone resin by vapor phase electron beam polymerization," *J.Vac.Sci.Technol.*, B1(4):1171-1173 (1983).
- Morrison et al., "Solution-Phase Detection of Polynucleotides Using Interacting Fluorescent Labels and Competitive Hybridization," *Anal. Biochem.*, 183:231-244 (1989).
- Munegumi et al., "thermal Synthesis of Polypeptides from N-Boc-Amino Acid (Aspartic Acid, β -Aminoglutamic Acid) Anhydrides," *Chem. Letters*, pp. 1643-1646 (1988).
- Mutter et al., "Impact of Conformation on the Synthetic Strategies for Peptide Sequences," pp. 217-228 from Chemistry of Peptides and Proteins, vol. 1, Proceedings of the Third USSR-FRG Symp., in USSR (1982).
- Nakamori et al., "A Simple and Useful Method for Simultaneous Screening of Elevated Levels of Expression of a Variety of Oncogenes in Malignant Cells," *Jpn. J. Cancer Res.*, 79:1311-1317 (1988).
- Nederlof et al., "Multiple Fluorescence In Situ Hybridization," *Cytometry*, 11:126-131 (1990).
- Nyborg, W., "Acoustic Streaming," chapter 11 pp. 265-329 from Physical Acoustics, Principles and Methods, Mason, eds., vol. II, part B, Academic Press, New York and London (1965).
- Oevirk et al., "High Performance Liquid Chromatography Partially Integrated onto a Silicon Chip," *Analyt. Meth. Instrumentation*, 2(2):74-82 (1995).

US 6,576,424 B2

Page 8

- Ohtsuka et al., "Studies on transfer ribonucleic acids and related compounds. IX Ribonucleic oligonucleotide synthesis using a photosensitive 0-nitrobenzyl protection at the 2' hydroxyl group," *Nuc.Acids.Res.*, 1(10):1351-1357 (1974).
- Olefirowicz et al., "Capillary Electrophoresis for Sampling Single Nerve Cells," *Chimia*, 45(4):106-108 (1991).
- Patchornik et al., "Photosensitive Protecting Groups," *J.Am.Chem.Soc.*, 92(21):6333-6335 (1970).
- Patent Abstracts of Japan from EPO, Abst. 13:557, JP 1-233 447 (1989).
- Pease et al., "Light-generated oligonucleotide arrays for rapid DNA sequence analysis," *PNAS*, 91:5022-26 (1994).
- Pevzner, P.A., "I-Tuple DNA Sequencing: Computer Analysis," *J. Biomol. Struct. Dynam.*, 7(1):63-69 (1989).
- Pfahler et al., "Liquid Transport in Micron and Submicron Channels," *Sensors and Actuators*, A21-A23:431-4 (90).
- Pidgeon et al., "Immobilized Artificial Membrane Chromatography: Supports Composed of Membrane Lipids," *Anal. Biochem.*, 176:36-47 (89).
- Pillai, V.N., "Photoremovable Protecting Groups in Organic Synthesis," *Synthesis*, pp. 1-26 (1980).
- Pillai et al., "3-Nitro-4-Aminomethylbenzoylderivate von Polyethylenglykolen: Eine neue Klasse von Photosensitiven loslichen Polymeren Tragern zur Synthese von C-terminalen Peptidamiden," *Tetrah. Ltr.*, # 36 p. 3409-3412 (1979).
- Pillai et al., "Synthetic Hydrophilic Polymers, Biomedical and Chemical Applications," *Naturwissenschaften*, 68:558-566 (1981).
- Pirrung et al., "Proofing of Photolithographic DNA Synthesis with 3' .5'-Dimethoxybenzoinyloxycarbonyl-Protected Deoxynucleoside Phosphoramidites," *J. Org. Chem.*, 63(2):241-246 (1998).
- Pirrung et al., "Comparison of Methods for Photochemical Phosphoramidite-Based DNA Synthesis," *J. Org. Chem.*, 60:6270-6276 (1995).
- Ploax et al., "Cyclization of peptides on a solid support," *Int. J. Peptide Protein Research*, 29:162-169 (1987).
- Polsky-Cynkin et al., "Use of DNA Immobilized on Plastic and Agarose Supports to Detect DNA by Sandwich Hybridization," *Clin. Chem.*, 31(9):1428-1443 (1985).
- Poustka et al., "Molecular Approaches to Mammalian Genetics," Cold Spring Harbor Symposia on Quantitative Biology, 51:131-139 (1986).
- Purushothaman et al., "Synthesis of 4,5-diarylimidazoline-2-thiones and their photoconversion to bis(4,5-diarylimidazol-2-yl) sulphides," *Ind. J. Chem.*, 29B:18-21 (1990).
- Quesada et al., "High-Sensitivity DNA Detection with a Laser-Exited Confocal Fluorescence Gel Scanner," *Biotechniques*, 10:616 (1991).
- Reichmanis et al., *J. Polymer Sci. Polymer Chem. Edition*, 23:1-8 (1985).
- Richter et al., "An Electrohydrodynamic Micropump," *IEEE*, pp. 99-104 (1990).
- Richter et al., "Electrohydrodynamic Pumping and Flow Measurement," *IEEE*, pp. 271-276 (1991).
- Richter et al., "A Micromachined electrohydrodynamic (EHD) pump," *Sensors and Actuators*, A29:159-168 (91).
- Robertson et al., "A General and Efficient Route for Chemical Aminoacylation of Transfer RNAs," *J. Am. Chem. Soc.*, 113:2722-2729 (1991).
- Rodda et al., "The Antibody Response to Myoglobin-I. Systematic Synthesis of Myoglobin Peptides Reveals Location and Substructure of Species-Dependent Continuous Antigenic Determinants," *Mol. Immunol.*, 23(6):603-610 (1986).
- Rodgers, R.P., "Data Processing of Immunoassay Results," *Manual of Clin. Lab. Immunol.*, 3rd ed., ch. 15, pp. 82-87 (1986).
- Rose, D.J., "Free-solution reactor for post-column fluorescence detection in capillary zone electrophoresis," *J. Chromatography*, 540:343-353 (1991).
- Rovero et al., "Synthesis of Cyclic Peptides on solid Support," *Tetrahed. Letters*, 32(23):2639-2642 (1991).
- Sambrook, *Molecular Cloning—A Laboratory Manual*, publ. in 1989 (not included).
- Saiki et al., "Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes," *PNAS*, 86:6230-6234 (1989).
- Saiki et al., "Analysis of enzymatically amplified β -globin and HLA-DQ α DNA with Allele-specific oligonucleotide probes," *Nature*, 324:163-166 (1986).
- Scharf et al., "HLA class II allelic variation and susceptibility to pemphigus vulgaris," *PNAS*, 85(10):3504-3508 (1988).
- Schuup et al., "Mechanistic Studies of the Photorearrangement of o-Nitrobenzyl Esters," *J. Photochem.*, 36:85-97 (1987).
- Seiler et al., "Planar Glass Chips for Capillary Electrophoresis: Repetitive Sample Injection, Quantitation, and Separation Efficiency," *Anal. Chem.*, 65:1481-1488 (1993).
- Seller et al., "Electroosmotic Pumping and Valveless Control of Fluid Flow within a Manifold of Capillaries on a Glass Chip," *Anal. Chem.*, 66:3485-3491 (1994).
- Semmelhack et al., "Selective Removal of Protecting Groups Using Controlled Potential Electrolysis," *J. Am. Chem. Society*, 94(14):5139-5140 (1972).
- Sheldon et al., "Matrix DNA Hybridization," *Clinical Chemistry*, 39(4):718-719 (1993).
- Shin et al., "Dehydrooligonopeptides. XI. Facile Synthesis of Various Kinds of Dehydrodi- and tripeptides, and Dehydroenkephalins Containing Tyr Residue by Using N-Carboxydehydrotyrosine Anhydride," *Bull. Chem. Soc. Jpn.*, 62:1127-1135 (1989).
- Sim et al., "Use of a cDNA Library for Studies on Evolution and Developmental Expression of the Chorion Multigene Families," *Cell*, 18:1303-1316 (1979).
- Smith et al., "A Novel Method for Delineating Antigenic Determinants: Peptide Synthesis and Radioimmunoassay Using the Same Solid Support," *Immunochemistry*, 14:565-568 (1977).
- Southern et al., "Report on the Sequencing by Hybridization Workshop," *Genomics*, 13:1378-1383 (1992).
- Southern et al., "Oligonucleotide hybridisations on glass supports: a novel linker for oligonucleotide synthesis and hybridization properties of oligonucleotides synthesized in situ," *Nuc. Acids Res.*, 20(7):1679-1684 (1992).
- Southern et al., "Analyzing and Comparing Nucleic Acid Sequences by Hybridization to Arrays of Oligonucleotides: Evaluation Using Experimental Models," *Genomics*, 13:1008-10017 (1992).
- Stemme et al., "A valveless diffuserr/nozzle-based fluid pump," *Sensors and Actuators*, A39:159-167 (1993).

US 6,576,424 B2

Page 9

- Stryer, L., "DNA Probes and Genes Can be Synthesized by Automated Solid-Phase Methods," from *Biochemistry*, Third Edition, published by W.H. Freeman & Co., (1988).
- Stuber et al., "Synthesis and photolytic cleavage of bovine insulin B22-30 on a nitrobenzoylglycyl-poly (ethylene glycol) support," *Int. J. Peptide Protein Res.*, 22(3):277-283 (1984).
- Sundberg et al., "Spatially-Addressable Immobilization of Macromolecules on Solid Supports," *J. Am. Chem. Soc.*, 117(49):12050-12057 (1995).
- Swedberg, S.A., "Use of non-ionic and zwitterionic surfactants to enhance selectivity in high-performance capillary electrophoresis, An apparent micellar electrokinetic capillary chromatography mechanism," *J. Chromatography*, 503:449-452 (1990).
- Titus et al., "Texas Red, a Hydrophilic, red-emitting fluorophore for use with fluorescein in dual parameter flow microfluorometric and fluorescence microscopic studies," *J. Immunol. Meth.*, 50:193-204 (1982).
- Tkachuk et al., "Detection of bcr-abl Fusion in chronic Myelogenous Leukemia by in situ Hybridization," *Science*, 250:559-562 (90).
- Trzeciak et al., "Synthesis of 'Head-to-Tail' Cyclized Peptides on Solid Support by Fmoc Chemistry," *Tetrahed. Letters*, 33(32):4557-4560 (1992).
- Tsien et al., "Control of Cytoplasmic Calcium with Photolabile Tetracarboxylate 2-Nitrobenzhydryl Chelators," *Biophys. J.*, 50:843-853 (1986).
- Tsutsumi et al., "Expression of L- and M- Type Pyruvate Kinase in Human Tissues," *Genomics*, 2:86-89 (1988).
- Turchinskii et al., "Multiple Hybridization in Genome Analysis, Reaction of Diamines and Bisulfate with Cytosine for Introduction of Nonradioactive labels Into DNA," *Molecular Biology*, 22:1229-1235 (1988).
- Turner et al., "Photochemical Activation of Acylated α -Thrombin," *J. Am. Chem. Soc.*, 109:1274-1275 (1987).
- Urdea et al., "A novel method for the rapid detection of specific nucleotide sequences in crude biological samples without blotting or radioactivity; application to the analysis of hepatitis B virus in human serum," *Gene*, 61:253-264 (1987).
- Urdea et al., "A comparison of non-radioisotopic hybridization assay methods using fluorescent, chemiluminescent and enzyme labeled synthetic oligodeoxyribonucleotide probes," *Nuc. Acids Res.*, 16(11):4937-4956 (1988).
- Van der Voort et al., "Design and Use of a Computer Controlled Confocal Microscope for Biological Applications," *Scanning*, 7(2):66-78 (1985).
- Van Hijfte et al., "Intramolecular 1,3-Diyl Trapping Reactions. A Formal Total Synthesis of -Coriolin," *J. Organic Chemistry*, 50:3942-3944 (1985).
- Veldkamp, W.B., "Binary optics: the optics technology of the 1990s," *CLEO 90*, vol. 7, paper # CMG6 (1990).
- Verlaan-de Vries et al., "A dot-blot screening procedure for mutated ras oncogenes using synthetic oligodeoxynucleotides," *Gene*, 50:313-320 (1986).
- Verpoorte et al., "Three-dimensional micro flow manifolds for miniaturized chemical analysis systems," *J. Micromech. Microeng.*, 4:246-256 (1994).
- Volkmut et al., "DNA electrophoresis in microlithographic arrays," *Nature*, 358:600-602 (1992).
- Voss et al., "The immobilization of oligonucleotides and their hybridization properties," *Biochem. Soc. Transact.*, 16:216-217 (1988).
- Walker et al., "Photolabile Protecting Groups for an Acetylcholine Receptor Ligand. Synthesis and Photochemistry of a New Class of o-Nitrobenzyl Derivatives and their Effects on Receptor Function," *Biochemistry*, 25:1799-1805 (1986).
- Wallace et al., "Hybridization of synthetic oligodeoxyribonucleotides to $\Phi\chi$ 174 DNA: the effect of single base pair mismatch," *Nuc. Acids Res.*, 11(6):3543-3557 (1979).
- Washizu et al., "Handling Biological Cells Using a Fluid Integrated Circuit," *IEEE Transactions Industry Applications*, 26(2):352-358 (1990).
- Werner et al., "Size-Dependent Separation of Proteins Denatured in SDS by Capillary Electrophoresis Using a Replaceable Sieving Matrix," *Anal. Biochem.*, 212:253-258 (1993).
- White et al., "An Evaluation of Confocal Versus Conventional Imaging of Biological Structures by Fluorescence Light Microscopy," *J. Cell Biol.*, 105(1):41-48 (1987).
- Widacki et al., "Biochemical Differences in Qa-2 Antigens Expressed by Qa-2+,6+ and Qa-2a+,6- Strains. Evidence for Differential Expression of the Q7 and Q9 Genes," *Mol. Immunology*, 27(6):559-570 (1990).
- Wilcox et al., "Synthesis of Photolabile 'Precursors' of Amino Acid Neurotransmitters," *J. Org. Chem.*, 55:1585-1589 (1990).
- Wilding et al., "PCR in a Silicon Microstructure," *Clin. Chem.*, 40(9):1815-1818 (1994).
- Wilding et al., "Manipulation and Flow of Biological Fluids in Straight Channels Micromachined in Silicon," *Clin. Chem.*, 40(1):43-47 (1994).
- Wittman-Liebold, eds., *Methods in Protein Sequence Analysis*, from Proceedings of 7th Int'l Conf., Berlin, Germany, 7/3-8/88, table of contents, pp. xi-xx* (1989).
- Woolley et al., "Ultra-high-speed DNA fragment separations using microfabricated capillary array electrophoresis chips," *PNAS*, 91:11348-11352 (1994).
- Wu et al., "Synthesis and Properties of Adenosine-5'-triphospho- γ -5-(5-sulfonic acid)naphthyl Ethylamide: A Fluorescent Nucleotide Substrate for DNA-Dependent RNA Polymerase from *Escherichia coli*," *Arch. Biochem. Biophys.*, 246(2):564-571 (1986).
- Wu et al., "Laboratory Methods, Direct Analysis of Single Nucleotide Variation in Human DNA and RNA Using In Situ Dot Hybridization," *DNA*, 8(2):135-142 (1989).
- Yamamoto et al., "Features and applications of the laser scanning microscope," *J. Mod. Optics*, 37(11):1691-1701 (1990).
- Yarbrough et al., "Synthesis and Properties of Fluorescent Nucleotide Substrates for DNA-dependent RNA Polymerases," *J. Biol. Chem.*, 254(23):12069-12073 (1979).
- Yosomiya et al., "Performance, Glass fiber Having Isocyanate Group on the Surface. Preparation and Reaction with Amino Acid," *Polymer Bulletin*, 12:41-48 (1984).
- Young, W.S., "Simultaneous Use of Digoxigenin- and Radiolabeled Oligodeoxyribonucleotide Probes for Hybridization Histochemistry," *Neuropeptides*, 13:271-275 (1989).
- Yue et al., "Miniature Field-Flow Fractionation System for Analysis of Blood Cells," *Clin. Chem.*, 40(9):1810-1814 (1994).
- Zehavi et al., "Light-Sensitive Glycosides. 1. 6-Nitroveratryl β -D-Glucopyranoside and 2-Nitrobenzyl β -D-Glucopyranoside," *J. Org. Chem.*, 37(14):2281-2285 (1972).

US 6,576,424 B2

Page 10

- Zengerle et al., "Transient measurements on miniaturized diaphragm pumps in microfluid systems," *Sensors and Actuators*, A46-47:557-561 (1995).
- Perkin Elmer Cetus, Gene Amp DNA Amplification Reagent Kit, insert, Oct. 1988.
- Church et al, Proc. Natl. Acad. Sci., 81:1991-1995 (Apr., 1984).
- Chetverin et al, Bio/Technology, 12:1093-1099 (Nov. 1994).
- Coulson et al, Proc. Natl. Acad. Sci. USA, 83:7821-7825 (Oct. 1986).
- Dower et al, Ann. Rep. Med. Chem., 26:271-280 (1991).
- Dramanac et al, J. Biomol. Struct. Dyn., 8(5):1085-1102 (1991).
- Hodgson et al, Nucl. Acids Res., 15(15):6295 (1987).
- Khrapko et al, DNA Seq. Map, 1:375-388 (1991).
- Lander et al, Genomics, 2:231-239 (1988).
- Little, Nature, 346:611-612 (1990).
- Lysov et al, Dokl. Akad. Nauk. SSSR, 303:1508-1511 (1988).
- Olson et al, Proc. Natl. Acad. Sci. USA, 83:7826-7830 (Oct. 1986).
- Pevzner, Algorithmica, 13(1-2):77-105 (1995).
- Pevzner et al, Algorithmica, 13(1-2):135-154 (1995).
- Pfeifer et al, Science, 246:810-813 (Nov. 10, 19889).
- Seed, Nucl. Acids Res., 10(5):1799-1810 (1982).
- Wood et al, Proc. Natl. Acad. Sci. USA, 82:1585-1588 (1985).
- Feinberg et al, Anal. Biochem., 137:266-267 (1984).
- Pevzner et al, Adv. Applied Math, 14:139-171 (1993).
- Schena et al, Proc. Natl. Acad. Sci. USA, 93:10614-10619 (Oct. 1996).
- Bannwarth "Gene technology: A challenge for a chemist" *Chimia* 1987, 41:302-317.
- Bannwarth et al. "A system for the simultaneous chemical synthesis of different DNA fragments on solid support" *DNA* 1986, 5:413-419.
- Brenner et al. "In vitro cloning of complex mixtures of DNA on microbeads: Physical separation of differentially expressed cDNAs" *Proc Natl Acad Sci USA* 2000, 97:1665-1670.
- Brenner et al. "Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays" *Nature Biotechnology* 2000, 18:630-634.
- Tyagi "Taking a census of mRNA populations with microbeads" *Nature Biotechnology* 2000, 18:597-598.
- Wada (chairman) *Hayashibara Intl Workshop on Automatic and High Speed DNA-Base Sequencing* 1987 pp. 1-63.
- Miller et al. "Detection of bacteria by hybridization of rRNA with DNA-latex and immunodetection of hybrids" *J Clin Microbiol* 1988, 26:1271-1276.

* cited by examiner

U.S. Patent

Jun. 10, 2003

Sheet 1 of 2

US 6,576,424 B2

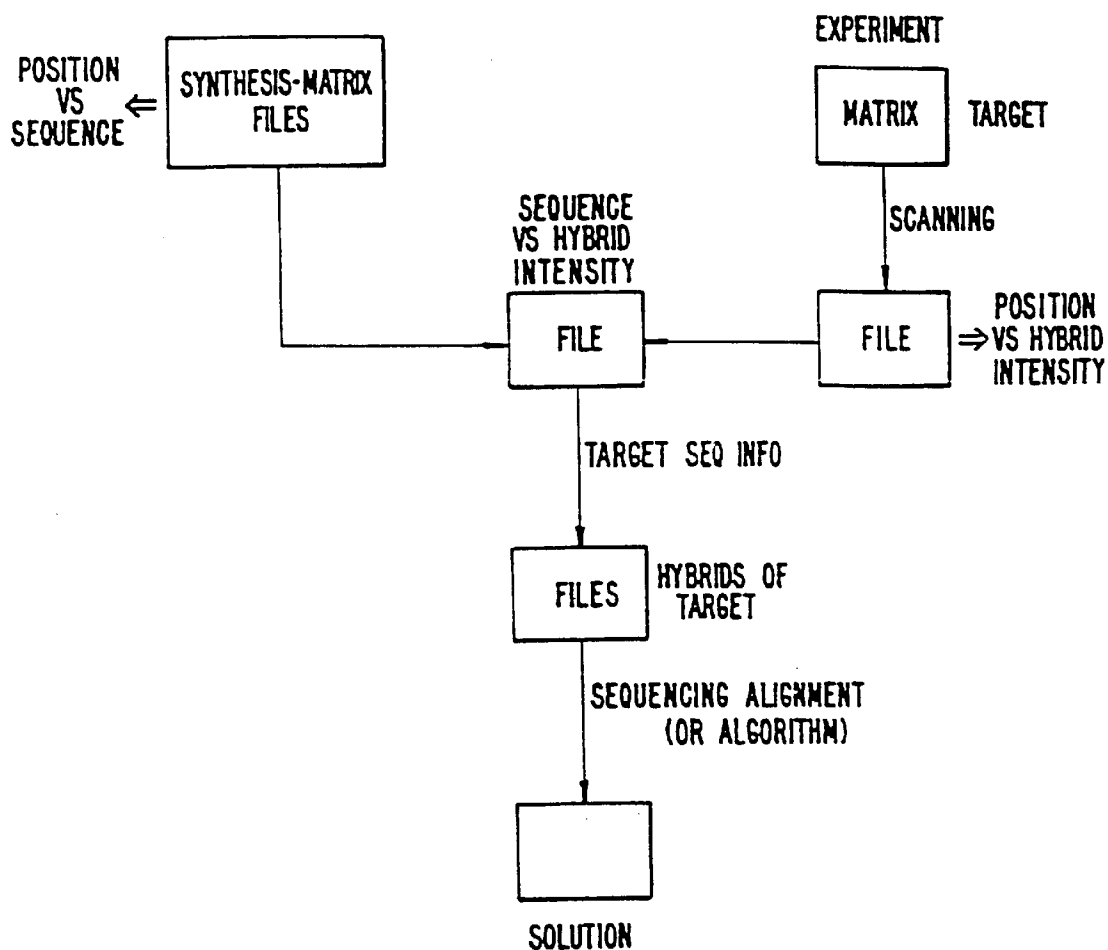


FIG. 1

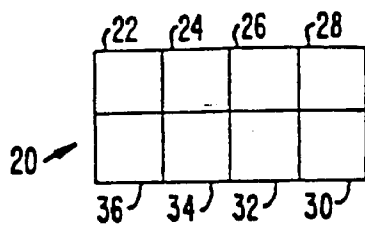


FIG. 2A

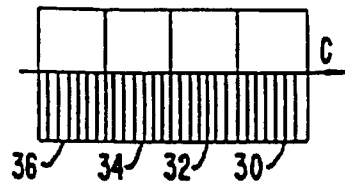


FIG. 2B

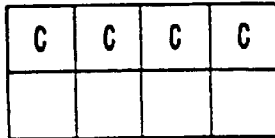


FIG. 2C

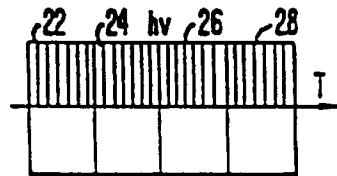


FIG. 2D

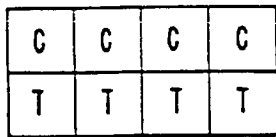


FIG. 2E

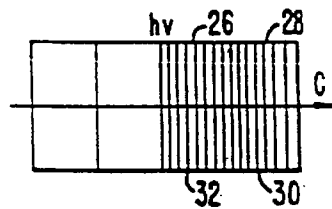


FIG. 2F

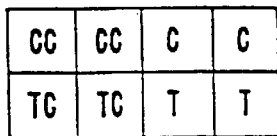


FIG. 2G

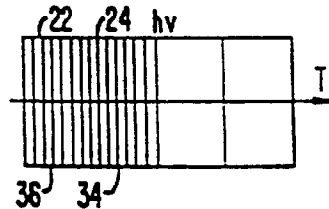


FIG. 2H

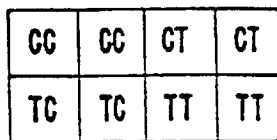


FIG. 2I

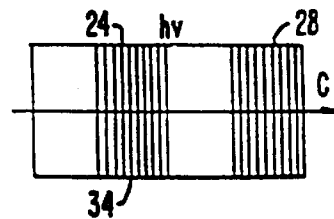


FIG. 2J

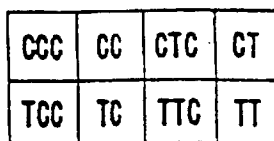


FIG. 2K

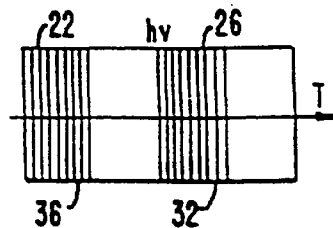


FIG. 2L

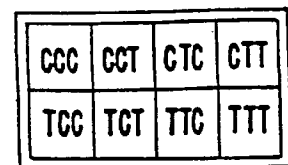


FIG. 2M

US 6,576,424 B2

1

ARRAYS AND METHODS FOR DETECTING NUCLEIC ACIDS

CROSS REFERENCE TO RELATED APPLICATIONS

This is a continuation of appln. Ser. No. 09/670,563, filed Sep. 27, 2000; which is a continuation of appln. Ser. No. 09/362,089, filed Jul. 28, 1999; which is a divisional of appln. Ser. No. 09/056,927, filed Apr. 8, 1998, now U.S. Pat. No. 6,197,506; which is a continuation of appln. Ser. No. 08/670,118, filed Jun. 25, 1996, now U.S. Pat. No. 5,800,992, which is a divisional of appln. Ser. No. 08/168,904, filed Dec. 15, 1993, now abandoned; which is a continuation of appln. Ser. No. 07/624,114, filed Dec. 6, 1990, now abandoned.

Additional commonly assigned appln Ser. No. 07/492,462, filed Mar. 7, 1990, now U.S. Pat. No. 5,143,854; and appln. Ser. No. 07/362,901, filed Jun. 7, 1989, now abandoned, as well as appln. Ser. Nos. 07/624,120 and 07/626,730, both filed on Dec. 6, 1990; appln. Ser. No. 07/435,316, filed Nov. 13, 1989, now abandoned; and U.S. Pat. No. 5,252,743 are hereby incorporated herein by reference.

BACKGROUND OF THE INVENTION

The present invention relates to the sequencing, fingerprinting, and mapping of polymers, particularly biological polymers. The inventions may be applied, for example, in the sequencing, fingerprinting, or mapping of nucleic acids, polypeptides, oligosaccharides, and synthetic polymers.

The relationship between structure and function of macromolecules is of fundamental importance in the understanding of biological systems. These relationships are important to understanding, for example, the functions of enzymes, structural proteins, and signalling proteins, ways in which cells communicate with each other, as well as mechanisms of cellular control and metabolic feedback.

Genetic information is critical in continuation of life processes. Life is substantially informationally based and its genetic content controls the growth and reproduction of the organism and its complements. Polypeptides, which are critical features of all living systems, are encoded by the genetic material of the cell. In particular, the properties of enzymes, functional proteins, and structural proteins are determined by the sequence of amino acids which make them up. As structure and function are integrally related, many biological functions may be explained by elucidating the underlying structural features which provide those functions. For this reason, it has become very important to determine the genetic sequences of nucleotides which encode the enzymes, structural proteins, and other effectors of biological functions. In addition to segments of nucleotides which encode polypeptides, there are many nucleotide sequences which are involved in control and regulation of gene expression.

The human genome project is directed toward determining the complete sequence of the genome of the human organism. Although such a sequence would not correspond to the sequence of any specific individual, it would provide significant information as to the general organization and specific sequences contained within segments from particular individuals. It would also provide mapping information which is very useful for further detailed studies. However, the need for highly rapid, accurate, and inexpensive sequencing technology is nowhere more apparent than in a

2

demanding sequencing project such as this. To complete the sequencing of a human genome would require the determination of approximately 3×10^9 , or 3 billion base pairs.

The procedures typically used today for sequencing include the Sanger dideoxy method, see, e.g., Sanger et al. (1977) *Proc. Natl. Acad. Sci. USA*, 74:5463-5467, or the Maxam and Gilbert method, see, e.g., Maxam et al., (1980) *Methods in Enzymology*, 65:499-559. The Sanger method utilizes enzymatic elongation procedures with chain terminating nucleotides. The Maxam and Gilbert method uses chemical reactions exhibiting specificity of reaction to generate nucleotide specific cleavages. Both methods require a practitioner to perform a large number of complex manual manipulations. These manipulations usually require isolating homogeneous DNA fragments, elaborate and tedious preparing of samples, preparing a separating gel, applying samples to the gel, electrophoresing the samples into this gel, working up the finished gel, and analyzing the results of the procedure.

Thus, a less expensive, highly reliable, and labor efficient means for sequencing biological macromolecules is needed. A substantial reduction in cost and increase in speed of nucleotide sequencing would be very much welcomed. In particular, an automated system would improve the reproducibility and accuracy of procedures. The present invention satisfies these and other needs.

SUMMARY OF THE INVENTION

The present invention provides improved methods useful for de novo sequencing of an unknown polymer sequence, for verification of known sequences, for fingerprinting polymers, and for mapping homologous segments within a sequence. By reducing the number of manual manipulations required and automating most of the steps, the speed, accuracy, and reliability of these procedures are greatly enhanced.

The production of a substrate having a matrix of positionally defined regions with attached reagents exhibiting known recognition specificity can be used for the sequence analysis of a polymer. Although most directly applicable to sequencing, the present invention is also applicable to fingerprinting, mapping, and general screening of specific interactions. The VLSIPS™ Technology (Very Large Scale Immobilized Polymer Synthesis) substrates will be applied to evaluating other polymers, e.g., carbohydrates, polypeptides, hydrocarbon synthetic polymers, and the like. For these non-polynucleotides, the sequence specific reagents will usually be antibodies specific for a particular subunit sequence.

According to one aspect of the masking technique, the invention provides an ordered method for forming a plurality of polymer sequences by sequential addition of reagents comprising the step of serially protecting and deprotecting portions of the plurality of polymer sequences for addition of other portions of the polymer sequences using a binary synthesis strategy.

The present invention also provides a means to automate sequencing manipulations. The automation of the substrate production method and of the scan and analysis steps minimizes the need for human intervention. This simplifies the tasks and promotes reproducibility.

The present invention provides a composition comprising a plurality of positionally distinguishable sequence specific reagents attached to a solid substrate, which reagents are capable of specifically binding to a predetermined subunit sequence of a preselected multi-subunit length having at

US 6,576,424 B2

3

least three subunits, said reagents representing substantially all possible sequences of said preselected length. In some embodiments, the subunit sequence is a polynucleotide or a polypeptide, in others the preselected multi-subunit length is five subunits and the subunit sequence is a polynucleotide sequence. In other embodiments, the specific reagent is an oligonucleotide of at least about five nucleotides. Alternatively, the specific reagent is a monoclonal antibody. Usually the specific reagents are all attached to a single solid substrate, and the reagents comprise about 3000 different sequences. In other embodiments, the reagents represents at least about 25% of the possible subsequences of said preselected length. Usually, the reagents are localized in regions of the substrate having a density of at least 25 regions per square centimeter, and often the substrate has a surface area of less than about 4 square centimeters.

The present invention also provides methods for analyzing a sequence of a polynucleotide or a polypeptide, said method comprising the step of:

- a) exposing said polynucleotide or polypeptide to a composition as described.

It also provides useful methods for identifying or comparing a target sequence with a reference, said method comprising the step of:

- a) exposing said target sequence to a composition as described;
- b) determining the pattern of positions of the reagents which specifically interact with the target sequence; and
- c) comparing the pattern with the pattern exhibited by the reference when exposed to the composition.

The present invention also provides methods for sequencing a segment of a polynucleotide comprising the steps of:

- a) combining:
 - i) a substrate comprising a plurality of chemically synthesized and positionally distinguishable oligonucleotides capable of recognizing defined oligonucleotide sequences; and
 - ii) a target polynucleotide; thereby forming high fidelity matched duplex structures of complementary subsequences of known sequence; and
- b) determining which of said reagents have specifically interacted with subsequences in said target polynucleotide.

In one embodiment, the segment is substantially the entire length of said polynucleotide.

The invention also provides methods for sequencing a polymer, said method comprising the steps of:

- a) preparing a plurality of reagents which each specifically bind to a subsequence of preselected length;
- b) positionally attaching each of said reagents to one or more solid phase substrates, thereby producing substrates of positionally definable sequence specific probes;
- c) combining said substrates with a target polymer whose sequence is to be determined; and
- d) determining which of said reagents have specifically interacted with subsequences in said target polymer.

In one embodiment, the substrates are beads. Preferably, the plurality of reagents comprise substantially all possible subsequences of said preselected length found in said target. In another embodiment, the solid phase substrate is a single substrate having attached thereto reagents recognizing substantially all possible subsequences of preselected length found in said target.

4

In another embodiment, the method further comprises the step of analyzing a plurality of said recognized subsequences to assemble a sequence of said target polymer. In a bead embodiment, at least some of the plurality of substrates have one subsequence specific reagent attached thereto, and the substrates are coded to indicate the sequence specificity of said reagent.

The present invention also embraces a method of using a fluorescent nucleotide to detect interactions with oligonucleotide probes of known sequence, said method comprising:

- a) attaching said nucleotide to a target unknown polynucleotide sequence, and
- b) exposing said target polynucleotide sequence to a collection of positionally defined oligonucleotide probes of known sequences to determine the sequences of said probes which interact with said target.

In a further refinement, an additional step is included of:

- a) collating said known sequences to determine the overlaps of said known sequences to determine the sequence of said target sequence.

A method of mapping a plurality of sequences relative to one another is also provided, the method comprising:

- a) preparing a substrate having a plurality of positionally attached sequence specific probes;
- b) exposing each of said sequences to said substrate, thereby determining the patterns of interaction between said sequence specific probes and said sequences; and
- c) determining the relative locations of said sequence specific probe interactions on said sequences to determine the overlaps and order of said sequences.

In one refinement, the sequence specific probes are oligonucleotides, applicable to where the target sequences are nucleic acid sequences.

In the nucleic acid sequencing application, the steps of the sequencing process comprise:

- a) producing a matrix substrate having known positionally defined regions of known sequence specific oligonucleotide probes;
- b) hybridizing a target polynucleotide to the positions on the matrix so that each of the positions which contain oligonucleotide probes complementary to a sequence on the target hybridize to the target molecule;
- c) detecting which positions have bound the target, thereby determining sequences which are found on the target; and
- d) analyzing the known sequences contained in the target to determine sequence overlaps and assembling the sequence of the target therefrom.

The enablement of the sequencing process by hybridization is based in large part upon the ability to synthesize a large number (e.g., to virtually saturate) of the possible overlapping sequence segments and distinguishing those probes which hybridize with fidelity from those which have mismatched bases, and to analyze a highly complex pattern of hybridization results to determine the overlap regions.

The detecting of the positions which bind the target sequence would typically be through a fluorescent label on the target. Although a fluorescent label is probably most convenient, other sorts of labels, e.g., radioactive, enzyme linked, optically detectable, or spectroscopic labels may be used. Because the oligonucleotide probes are positionally defined, the location of the hybridized duplex will directly translate to the sequences which hybridize. Thus, analysis of the positions provides a collection of subsequences found within the target sequence. These subsequences are matched

US 6,576,424 B2

5

with respect to their overlaps so as to assemble an intact target sequence.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 illustrates a flow chart for sequence, fingerprint, or mapping analysis.

FIGS. 2A–M illustrates the process of a VLSIPS™ Technology trinucleotide synthesis.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

I. Overall Description

- A. general
- B. VLSIPS substrates
- C. binary masking
- D. applications
- E. detection methods and apparatus
- F. data analysis

II. Theoretical Analysis

- A. simple n-mer structure; theory
- B. complications
- C. non-polynucleotide embodiments

III. Polynucleotide Sequencing

- A. preparation of substrate matrix
- B. labeling target polynucleotide
- C. hybridization conditions
- D. detection; VLSIPS scanning
- E. analysis
- F. substrate reuse
- G. non-polynucleotide aspects

IV. Fingerprinting

- A. general
- B. preparation of substrate matrix
- C. labeling target nucleotides
- D. hybridization conditions
- E. detection; VLSIPS scanning
- F. analysis
- G. substrate reuse
- H. non-polynucleotide aspects

V. Mapping

- A. general
- B. preparation of substrate matrix
- C. labeling
- D. hybridization/specific interaction
- E. detection
- F. analysis
- G. substrate reuse
- H. non-polynucleotide aspects

VI. Additional Screening

- A. specific interactions
- B. sequence comparisons
- C. categorizations
- D. statistical correlations

VII. Formation of Substrate

- A. instrumentation
- B. binary masking
- C. synthetic methods
- D. surface immobilization

6

VIII. Hybridization/Specific Interaction

- A. general
- B. important parameters

IX. Detection Methods

- A. labeling techniques
- B. scanning system

X. Data Analysis

- A. general
- B. hardware
- C. software

XI. Substrate Reuse

- A. removal of label
- B. storage and preservation

XII. Integrated Sequencing Strategy

- A. initial mapping strategy
- B. selection of smaller clones
- C. actual sequencing procedures

XIII. Commercial Applications

- A. sequencing
- B. fingerprinting
- C. mapping

I. OVERALL DESCRIPTION

A. General

The present invention relies in part on the ability to synthesize or attach specific recognition reagents at known locations on a substrate, typically a single substrate. In particular, the present invention provides the ability to prepare a substrate having a very high density matrix pattern of positionally defined specific recognition reagents. The reagents are capable of interacting with their specific targets while attached to the substrate, e.g., solid phase interactions, and by appropriate labeling of these targets, the sites of the interactions between the target and the specific reagents may be derived. Because the reagents are positionally defined, the sites of the interactions will define the specificity of each interaction. As a result, a map of the patterns of interactions with specific reagents on the substrate is convertible into information on the specific interactions taking place, e.g., the recognized features. Where the specific reagents recognize a large number of possible features, this system allows the determination of the combination of specific interactions which exist on the target molecule. Where the number of features is sufficiently large, the identical same combination, or pattern, of features is sufficiently unlikely that a particular target molecule may often be uniquely defined by its features. In the extreme, the features may actually be the subunit sequence of the target molecule, and a given target sequence may be uniquely defined by its combination of features.

In particular, the methodology is applicable to sequencing polynucleotides. The specific sequence recognition reagents will typically be oligonucleotide probes which hybridize with specificity to subsequences found on the target sequence. A sufficiently large number of those probes allows the fingerprinting of a target polynucleotide or the relative mapping of a collection of target polynucleotides, as described in greater detail below.

In the high resolution fingerprinting provided by a saturating collection of probes which include all possible subsequences of a given size, e.g., 10-mers, collating of all the subsequences and determination of specific overlaps will be derived and the entire sequence can usually be reconstructed.

US 6,576,424 B2

7

Although a polynucleotide sequence analysis is a preferred embodiment, for which the specific reagents are most easily accessible, the invention is also applicable to analysis of other polymers, including polypeptides, carbohydrates, and synthetic polymers, including α -, β -, and ω -amino acids, polyurethanes, polyesters, polycarbonates, polyureas, polyamides, polyethyleneimines, polyarylene sulfides, polysiloxanes, polyimides, polyacetates, and mixed polymers. Various optical isomers, e.g., various D- and L-forms of the monomers, may be used.

Sequence analysis will take the form of complete sequence determination, to the level of the sequence of individual subunits along the entire length of the target sequence. Sequence analysis also takes the form of sequence homology, e.g., less than absolute subunit resolution, where "similarity" in the sequence will be detectable, or the form of selective sequences of homology interspersed at specific or irregular locations.

In either case, the sequence is determinable at selective resolution or at particular locations. Thus, the hybridization method will be useful as a means for identification, e.g., a "fingerprint", much like a Southern hybridization method is used. It is also useful to map particular target sequences.

B. VLSIPS™ Technology

The invention is enabled by the development of technology to prepare substrates on which specific reagents may be either positionally attached or synthesized. In particular, the very large scale immobilized polymer synthesis (VLSIPS™) technology allows for the very high density production of an enormous diversity of reagents mapped out in a known matrix pattern on a substrate. These reagents specifically recognize subsequences in a target polymer and bind thereto, producing a map of positionally defined regions of interaction. These map positions are convertible into actual features recognized, and thus would be present in the target molecule of interest.

As indicated, the sequence specific recognition reagents will often be oligonucleotides which hybridize with fidelity and discrimination to the target sequence. For use with other polymers, monoclonal or polyclonal antibodies having high sequence specificity will often be used.

In the generic sense, the VLSIPS technology allows the production of a substrate with a high density matrix of positionally mapped regions with specific recognition reagents attached at each distinct region. By use of protective groups which can be positionally removed, or added, the regions can be activated or deactivated for addition of particular reagents or compounds. Details of the protection are described below and in related Pirrung et al. (1992) U.S. Pat. No. 5,143,854. In a preferred embodiment, photosensitive protecting agents will be used and the regions of activation or deactivation may be controlled by electro-optical and optical methods, similar to many of the processes used in semiconductor wafer and chip fabrication.

In the nucleic acid nucleotide sequencing application, a VLSIPS substrate is synthesized having positionally defined oligonucleotide probes. See Pirrung et al. (1992) U.S. Pat. No. 5,143,854; and U.S. Ser. No. 07/624,120, now abandoned. By use of masking technology and photosensitive synthetic subunits, the VLSIPS apparatus allows for the stepwise synthesis of polymers according to a positionally defined matrix pattern. Each oligonucleotide probe will be synthesized at known and defined positional locations on the substrate. This forms a matrix pattern of known relationship between position and specificity of interaction. The VLSIPS technology allows the production of a very large number of

8

different oligonucleotide probes to be simultaneously and automatically synthesized including numbers in excess of about 10^2 , 10^3 , 10^4 , 10^5 , 10^6 , or even more, and at densities of at least about 10^2 , $10^3/\text{cm}^2$, $10^4/\text{cm}^2$, $10^5/\text{cm}^2$ and up to $10^6/\text{cm}^2$ or more. This application discloses methods for synthesizing polymers on a silicon or other suitably derivatized substrate, methods and chemistry for synthesizing specific types of biological polymers on those substrates, apparatus for scanning and detecting whether interaction has occurred at specific locations on the substrate, and various other technologies related to the use of a high density very large scale immobilized polymer substrate. In particular, sequencing, fingerprinting, and mapping applications are discussed herein in detail, though related technologies are described in simultaneously filed applications U.S. Ser. No. 07/624,120, now abandoned; and U.S. Ser. No. 07/517,659; Dower et al. (1995) U.S. Pat. No. 5,427,908, each of which is hereby incorporated herein by reference.

In other embodiments, antibody probes will be generated which specifically recognize particular subsequences found on a polymer. Antibodies would be generated which are specific for recognizing a three contiguous amino acid sequence, and monoclonal antibodies may be preferred. Optimally, these antibodies would not recognize any sequences other than the specific three amino acid stretch desired and the binding affinity should be insensitive to flanking or remote sequences found on a target molecule. Likewise, antibodies specific for particular carbohydrate linkages or sequences will be generated. A similar approach could be used for preparing specific reagents which recognize other polymer subunit sequences. These reagents would typically be site specifically localized to a substrate matrix pattern where the regions are closely packed.

These reagents could be individually attached at specific sites on the substrate in a matrix by an automated procedure where the regions are positionally targeted by some other specific mechanism, e.g., one which would allow the entire collection of reagents to be attached to the substrate in a single reaction. Each reagent could be separately attached to a specific oligonucleotide sequence by an automated procedure. This would produce a collection of reagents where, e.g., each monoclonal antibody would have a unique oligonucleotide sequence attached to it. By virtue of a VLSIPS substrate which has different complementary oligonucleotides synthesized on it, each monoclonal antibody would specifically be bound only at that site on the substrate where the complementary oligonucleotide has been synthesized. A crosslinking step would fix the reagent to the substrate. See, e.g., Dattagupta et al. (1985) U.S. Pat. No. 4,542,102 and (1987) U.S. Pat. No. 4,713,326; and Chatterjee, M. et al. (1990) *J. Am. Chem. Soc.* 112:6397-6399, which are hereby incorporated herein by reference. This allows a high density positionally specific collection of specific recognition reagents, e.g., monoclonal antibodies, to be immobilized to a solid substrate using an automated system.

The regions which define particular reagents will usually be generated by selective protecting groups which may be activated or deactivated. Typically the protecting group will be bound to a monomer subunit or spatial region, and can be spatially affected by an activator, such as electromagnetic radiation. Examples of protective groups with utility herein include nitroveratryl oxycarbonyl (NVOC), nitrobenzyl oxycarbonyl (NBOC), dimethyl dimethoxy benzyloxy carbonyl, 5-bromo-7-nitroindolyl, O-hydroxy- α -methyl cinnamoyl, and 2-oxymethylene anthraquinone. Examples of activators include ion beams, electric fields, magnetic fields, electron beams, x-ray, and other forms of electromagnetic radiation.

US 6,576,424 B2

9

C. Binary Masking

In fact, the means for producing a substrate useful for these techniques are explained in Pirrung et al. (1992) U.S. Pat. No. 5,143,854, which is hereby incorporated herein by reference. However, there are various particular ways to optimize the synthetic processes. Many of these methods are described in Ser. No. 07/624,120, now abandoned.

Briefly, the binary synthesis strategy refers to an ordered strategy for parallel synthesis of diverse polymer sequences by sequential addition of reagents which may be represented by a reactant matrix, and a switch matrix, the product of which is a product matrix. A reactant matrix is a 1xn matrix of the building blocks to be added. The switch matrix is all or a subset of the binary numbers from 1 to n arranged in columns. In preferred embodiments, a binary strategy is one in which at least two successive steps illuminate half of a region of interest on the substrate. In most preferred embodiments, binary synthesis refers to a synthesis strategy which also factors a previous addition step. For example, a strategy in which a switch matrix for a masking strategy halves regions that were previously illuminated, illuminating about half of the previously illuminated region and protecting the remaining half (while also protecting about half of previously protected regions and illuminating about half of previously protected regions). It will be recognized that binary rounds may be interspersed with non-binary rounds and that only a portion of a substrate may be subjected to a binary scheme, but will still be considered to be a binary masking scheme within the definition herein. A binary "masking" strategy is a binary synthesis which uses light to remove protective groups from materials for addition of other materials such as nucleotides or amino acids.

In particular, this procedure provides a simplified and highly efficient method for saturating all possible sequences of a defined length polymer. This masking strategy is also particularly useful in producing all possible oligonucleotide sequence probes of a given length.

D. Applications

The technology provided by the present invention has very broad applications. Although described specifically for polynucleotide sequences, similar sequencing, fingerprinting, mapping, and screening procedures can be applied to polypeptide, carbohydrate, or other polymers. In particular, the present invention may be used to completely sequence a given target sequence to subunit resolution. This may be for de novo sequencing, or may be used in conjunction with a second sequencing procedure to provide independent verification. See, e.g., (1988) *Science* 242:1245. For example, a large polynucleotide sequence defined by either the Maxam and Gilbert technique or by the Sanger technique may be verified by using the present invention.

In addition, by selection of appropriate probes, a polynucleotide sequence can be fingerprinted. Fingerprinting is a less detailed sequence analysis which usually involves the characterization of a sequence by a combination of defined features. Sequence fingerprinting is particularly useful because the repertoire of possible features which can be tested is virtually infinite. Moreover, the stringency of matching is also variable depending upon the application. A Southern Blot analysis may be characterized as a means of simple fingerprint analysis.

Fingerprinting analysis may be performed to the resolution of specific nucleotides, or may be used to determine homologies, most commonly for large segments. In particular, an array of oligonucleotide probes of virtually any workable size may be positionally localized on a matrix

10

and used to probe a sequence for either absolute complementary matching, or homology to the desired level of stringency using selected hybridization conditions.

In addition, the present invention provides means for mapping analysis of a target sequence or sequences. Mapping will usually involve the sequential ordering of a plurality of various sequences, or may involve the localization of a particular sequence within a plurality of sequences. This may be achieved by immobilizing particular large segments onto the matrix and probing with a shorter sequence to determine which of the large sequences contain that smaller sequence. Alternatively, relatively shorter probes of known or random sequence may be immobilized to the matrix and a map of various different target sequences may be determined from overlaps. Principles of such an approach are described in some detail by Evans et al. (1989) "Physical Mapping of Complex Genomes by Cosmid Multiplex Analysis," *Proc. Natl. Acad. Sci. USA* 86:5030-5034; Michiels et al. (1987) "Molecular Approaches to Genome Analysis: A Strategy for the Construction of Ordered Overlap Clone Libraries," *CABIOS* 3:203-210; Olsen et al. (1986) "Random-Clone Strategy for Genomic Restriction Mapping in Yeast," *Proc. Natl. Acad. Sci. USA* 83:7826-7830; Craig, et al. (1990) "Ordering of Cosmid Clones Covering the Herpes Simplex Virus Type I (HSV-I) Genome: A Test Case for Fingerprinting by Hybridization," *Nuc. Acids Res.* 18:2653-2660; and Coulson, et al. (1986) "Toward a Physical Map of the Genome of the Nematode *Caenorhabditis elegans*," *Proc. Natl. Acad. Sci. USA* 83:7821-7825; each of which is hereby incorporated herein by reference.

Fingerprinting analysis also provides a means of identification. In addition to its value in apprehension of criminals from whom a biological sample, e.g., blood, has been collected, fingerprinting can ensure personal identification for other reasons. For example, it may be useful for identification of bodies in tragedies such as fire, flood, and vehicle crashes. In other cases the identification may be useful in identification of persons suffering from amnesia, or of missing persons. Other forensics applications include establishing the identity of a person, e.g., military identification "dog tags", or may be used in identifying the source of particular biological samples. Fingerprinting technology is described, e.g., in Carrano, et al. (1989) "A High-Resolution, Fluorescence-Based, Semi-automated method for DNA Fingerprinting," *Genomics* 4: 129-136, which is hereby incorporated herein by reference. See, e.g., table I, for nucleic acid applications, and corresponding applications may be accomplished using polypeptides.

TABLE I

VLSIPS™ TECHNOLOGY IN NUCLEIC ACIDS

I. Construction of Chips

II. Applications

A. Sequencing

1. Primary sequencing
2. Secondary sequencing (sequence checking)
3. Large scale mapping
4. Fingerprinting

B. Duplex/Triplex formation

1. Antisense
2. Sequence specific function modulation (e.g. promoter inhibition)

US 6,576,424 B2

11

- C. Diagnosis
 - 1. Genetic markers
 - 2. Type markers
 - a. Blood donors
 - b. Tissue transplants
- D. Microbiology
 - 1. Clinical microbiology
 - 2. Food microbiology
- III. Instrumentation
 - A. Chip machines
 - B. Detection
- IV. Software Development
 - A. Instrumentation software
 - B. Data reduction software
 - C. Sequence analysis software

The fingerprinting analysis may be used to perform various types of genetic screening. For example, a single substrate may be generated with a plurality of screening probes, allowing for the simultaneous genetic screening for a large number of genetic markers. Thus, prenatal or diagnostic screening can be simplified, economized, and made more generally accessible.

In addition to the sequencing, fingerprinting, and mapping applications, the present invention also provides means for determining specificity of interaction with particular sequences. Many of these applications were described in Ser. No. 07/362,901, now abandoned, Pirrung et al. (1992) U.S. Pat. No. 5,143,854; Ser. No. 07/435,316, and Ser. No. 07/612,671.

E. Detection Methods and Apparatus

An appropriate detection method applicable to the selected labeling method can be selected. Suitable labels include radionucleotides, enzymes, substrates, cofactors, inhibitors, magnetic particles, heavy metal atoms, and particularly fluorescers, chemiluminescers, and spectroscopic labels. Patents teaching the use of such labels include U.S. Pat. Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,366,241.

With an appropriate label selected, the detection system best adapted for high resolution and high sensitivity detection may be selected. As indicated above, an optically detectable system, e.g., fluorescence or chemiluminescence would be preferred. Other detection systems may be adapted to the purpose, e.g., electron microscopy, scanning electron microscopy (SEM), scanning tunneling electron microscopy (STEM), infrared microscopy, atomic force microscopy (AFM), electrical conductance, and image plate transfer.

With a detection method selected, an apparatus for scanning the substrate will be designed. Apparatus, as described in Ser. No. 07/362,901, now abandoned; or Pirrung et al. (1992) U.S. Pat. No. 5,143,854; or Ser. No. 07/624,120, now abandoned, are particularly appropriate. Design modifications may also be incorporated therein.

F. Data Analysis

Data is analyzed by processes similar to those described below in the section describing theoretical analysis. More efficient algorithms will be mathematically devised, and will usually be designed to be performed on a computer. Various computer programs which may more quickly or efficiently make measurement samples and distinguish signal from noise will also be devised. See, particularly, Ser. No. 07/624,120, now abandoned.

The initial data resulting from the detection system is an array of data indicative of fluorescent intensity versus location on the substrate. The data are typically taken over regions substantially smaller than the area in which synthe-

12

sis of a given polymer has taken place. Merely by way of example, if polymers were synthesized in squares on the substrate having dimensions of 500 microns by 500 microns, the data may be taken over regions having dimensions of 5 microns by 5 microns. In most preferred embodiments, the regions over which fluorescence data are taken across the substrate are less than about $\frac{1}{2}$ the area of the regions in which individual polymers are synthesized, preferably less than $\frac{1}{10}$ the area in which a single polymer is synthesized, and most preferably less than $\frac{1}{100}$ the area in which a single polymer is synthesized. Hence, within any area in which a given polymer has been synthesized, a large number of fluorescence data points are collected.

A plot of number of pixels versus intensity for a scan should bear a rough resemblance to a bell curve, but spurious data are observed, particularly at higher intensities. Since it is desirable to use an average of fluorescent intensity over a given synthesis region in determining relative binding affinity, these spurious data will tend to undesirably skew the data.

Accordingly, in one embodiment of the invention the data are corrected for removal of these spurious data points, and an average of the data points is thereafter utilized in determining relative binding efficiency. In general the data are fitted to a base curve and statistical measures are used to remove spurious data.

In an additional analytical tool, various degeneracy reducing analogues may be incorporated in the hybridization probes. Various aspects of this strategy are described, e.g., in Macevicz, S. (1990) PCT publication number WO 90/04652, which is hereby incorporated herein by reference.

II. THEORETICAL ANALYSIS

The principle of the hybridization sequencing procedure is based, in part, upon the ability to determine overlaps of short segments. The VLSIPS technology provides the ability to generate reagents which will saturate the possible short subsequence recognition possibilities. The principle is most easily illustrated by using a binary sequence, such as a sequence of zeros and ones. Once having illustrated the application to a binary alphabet, the principle may easily be understood to encompass three letter, four letter, five or more letter, even 20 letter alphabets. A theoretical treatment of analysis of subsequence information to reconstruction of a target sequence is provided, e.g., in Lysov, Yu., et al. (1988) *Doklady Akademi. Nauk. SSR* 303:1508-1511; Khrapko K., et al. (1989) *FEBS Letters* 256:118-122; Pevzner, P. (1989) *J. of Biomolecular Structure and Dynamics* 7:63-69; and Drmanac, R. et al. (1989) *Genomics* 4:114-128; each of which is hereby incorporated herein by reference.

The reagents for recognizing the subsequences will usually be specific for recognizing a particular polymer subsequence anywhere within a target polymer. It is preferable that conditions may be devised which allow absolute discrimination between high fidelity matching and very low levels of mismatching. The reagent interaction will preferably exhibit no sensitivity to flanking sequences, to the subsequence position within the target, or to any other remote structure within the sequence. For polynucleotide sequencing, the specific reagents can be oligonucleotide probes; for polypeptides and carbohydrates, antibodies will be useful reagents. Antibody reagents should also be useful for other types of polymers.

A. Simple n-mer Structure: Theory

1. Simple Two Letter Alphabet: Example

A simple example is presented below of how a sequence of ten digits comprising zeros and ones would be sequence-

US 6,576,424 B2

13

able using short segments of five digits. For example, consider the sample ten digit sequence:

1010011100.

A VLSIPS™ Technology substrate could be constructed, as discussed elsewhere, which would have reagents attached in a defined matrix pattern which specifically recognize each of the possible five digit sequences of ones and zeros. The number of possible five digit subsequences is $2^5=32$. The number of possible different sequences 10 digits long is $2^{10}=1,024$. The five contiguous digit subsequences within a ten digit sequence number six, i.e., positioned at digits 1–5, 2–6, 3–7, 4–8, 5–9, and 6–10. It will be noted that the specific order of the digits in the sequence is important and that the order is directional, e.g., running left to right versus right to left.

The first five digit sequence contained in the target sequence is 10100. The second is 01001, the third is 10011, the fourth is 00111, the fifth is 01110, and the sixth is 11100.

The VLSIPS™ substrate would have a matrix pattern of positionally attached reagents which recognize each of the different 5-mer subsequences. Those reagents which recognize each of the 6 contained 5-mers will bind the target, and a label allows the positional determination of where the sequence specific interaction has occurred. By correlation of the position in the matrix pattern, the corresponding bound subsequences can be determined.

In the above-mentioned sequence, six different 5-mer sequences would be determined to be present. They would be:

10100

01001

10011

00111

01110

11100

Any sequence which contains the first five digit sequence, 10100, already narrows the number of possible sequences (e.g., from 1024 possible sequences) which contain it to less than about 192 possible sequences.

This 192 is derived from the observation that with the subsequence 10100 at the far left of the sequence, in positions 1–5, there are only 32 possible sequences. Likewise, for that particular subsequence in positions 2–6, 3–7, 4–8, 5–9, and 6–10. So, to sum up all of the sequences that could contain 10100, there are 32 for each position and 6 positions for a total of about 192 possible sequences. However, some of these 10 digit sequences will have been counted twice. Thus, by virtue of containing the 10100 subsequence, the number of possible 10-mer sequences has been decreased from 1024 sequences to less than about 192 sequences.

In this example, not only do we know that the sequence contains 10100, but we also know that it contains the second five character sequence, 01001. By virtue of knowing that the sequence contains 10100, we can look specifically to determine whether the sequence contains a subsequence of five characters which contains the four leftmost digits plus a next digit to the left. For example, we would look for a sequence of X1010, but we find that there is none. Thus, we know that the 10100 must be at the left end of the 10-mer. We would also look to see whether the sequence contains the rightmost four digits plus a next digit to the right, e.g., 0100X. We find that the sequence also contains the sequence 01001, and that X is a 1. Thus, we know at least that our target sequence has an overlap of 0100 and has the left terminal sequence 101001.

Applying the same procedure to the second 5-mer, we also know that the sequence must include a sequence of five

14

digits having the sequence 1001Y where Y must be either 0 or 1. We look through the fragments and we see that we have a 10011 sequence within our target, thus Y is also 1. Thus, we would know that our sequence has a sequence of the first seven being 1010011.

Moving to the next 5-mer, we know that there must be a sequence of 0011Z, where Z must be either 0 or 1. We look at the fragments produced above and see that the target sequence contains a 00111 subsequence and Z is 1. Thus, we know the sequence must start with 10100111.

The next 5-mer must be of the sequence 0111W where W must be 0 or 1. Again, looking up at the fragments produced, we see that the target sequence contains a 01110 subsequence, and W is a 0. Thus, our sequence to this point is 101001110. We know that the last 5-mer must be either 11100 or 11101. Looking above, we see that it is 11100 and that must be the last of our sequence. Thus, we have determined that our sequence must have been 1010011100.

However, it will be recognized from the example above with the sequences provided therein, that the sequence analysis can start with any known positive probe subsequence. The determination may be performed by moving linearly along the sequence checking the known sequence with a limited number of next positions. Given this possibility, the sequence may be determined, besides by scanning all possible oligonucleotide probe positions, by specifically looking only where the next possible positions would be. This may increase the complexity of the scanning but may provide a longer time span dedicated towards scanning and detecting specific positions of interest relative to other sequence possibilities. Thus, the scanning apparatus could be set up to work its way along a sequence from a given contained oligonucleotide to only look at those positions on the substrate which are expected to have a positive signal.

It is seen that given a sequence, it can be deconstructed into n-mers to produce a set of internal contiguous subsequences. From any given target sequence, we would be able to determine what fragments would result. The hybridization sequence method depends, in part, upon being able to work in the reverse, from a set of fragments of known sequences to the full sequence. In simple cases, one is able to start at a single position and work in either or both directions towards the ends of the sequence as illustrated in the example.

The number of possible sequences of a given length increases very quickly with the length of that sequence. Thus, a 10-mer of zeros and ones has 1024 possibilities, a 12-mer has 4096. A 20-mer has over a million possibilities, and a 30-mer has over a billion. However, a given 30-mer has, at most, 26 different internal 5-mer sequences. Thus, a 30 character target sequence having over a million possible sequences can be substantially defined by only 26 different 5-mers. It will be recognized that the probe oligonucleotides will preferably, but need not necessarily, be of identical length, and that the probe sequences need not necessarily be contiguous in that the overlapping subsequences need not differ by only a single subunit. Moreover, each position of the matrix pattern need not be homogeneous, but may actually contain a plurality of probes of known sequence. In addition, although all of the possible subsequence specifications would be preferred, a less than full set of sequences specifications could be used. In particular, although a substantial fraction will preferably be at least about 70%, it may be less than that. About 20% would be preferred, more preferably at least about 30% would be desired. Higher percentages would be especially preferred.

US 6,576,424 B2

15

2. Example of Four Letter Alphabet

A four letter alphabet may be conceptualized in at least two different ways from the two letter alphabet. One way is to consider the four possible values at each position and to analogize in a similar fashion to the binary example each of the overlaps. A second way is to group the binary digits into groups.

Using the first means, the overlap comparisons are performed with a four letter alphabet rather than a two letter alphabet. Then, in contrast to the binary system with 10 positions where $2^{10}=1024$ possible sequences, in a 4-character alphabet with 10 positions, there will actually be $4^{10}=1,048,576$ possible sequences. Thus, the complexity of a four character sequence has a much larger number of possible sequences compared to a two character sequence. Note, however, that there are still only 6 different internal 5-mers. For simplicity, we shall examine a 5 character string with 3 character subsequences. Instead of only 1 and 0, the characters may be designated, e.g., A, C, G, and T. Let us take the sequence GGCTA. The 3-mer subsequences are:

GCT

CTA

Given these subsequences, there is one sequence, or at most only a few sequences which would produce that combination of subsequences, i.e., GGCTA.

Alternatively, with a four character universe, the binary system can be looked at in pairs of digits. The pairs would be 00, 01, 10, and 11. In this manner, the earlier used sequence 1010011100 is looked at as 10,10,01,11,00. Then the first character of two digits is selected from the possible universe of the four representations 00, 01, 10, and 11. Then a probe would be in an even number of digits, e.g., not five digits, but, three pairs of digits or six digits. A similar comparison is performed and the possible overlaps determined. The 3-pair subsequences are:

10,10,01

10,01,11

01,11,00

and the overlap reconstruction produces 10,10,01,11,00.

The latter of the two conceptual views of the 4 letter alphabet provides a representation which is similar to what would be provided in a digital computer. The applicability to a four nucleotide alphabet is easily seen by assigning, e.g., 00 to A, 01 to C, 10 to G, and 11 to T. And, in fact, if such a correspondence is used, both examples for the 4 character sequences can be seen to represent the same target sequence. The applicability of the hybridization method and its analysis for determining the ultimate sequence is easily seen if A is the representation of adenine, C is the representation of cytosine, G is the representation of guanine, and T is the representation of thymine or uracil.

3. Generalization to m-Letter Alphabet

This reconstruction process may be applied to polymers of virtually any number of possible characters in the alphabet, and for virtually any length sequence to be sequenced, though limitations, as discussed below, will limit its efficiency at various extremes of length. It will be recognized that the theory can be applied to a large diversity of systems where sequence is important.

For example, the method could be applied to sequencing of a polypeptide. A polypeptide can have any of twenty natural amino acid possibilities at each position. A twenty letter alphabet is amenable to sequencing by this method so long as reagents exist for recognizing shorter subsequences therein. A preferred reagent for achieving that goal would be a set of monoclonal antibodies each of which recognizes a specific three contiguous amino acid subsequence. A com-

16

plete set of antibodies which recognize all possible subsequences of a given length, e.g., 3 amino acids, and preferably with a uniform affinity, would be $20^3=8000$ reagents.

It will also be recognized that each target sequence which is recognized by the specific reagents need not have homogeneous termini. Thus, fragments of the entire target sequence will also be useful for hybridizing appropriate subsequences. It is, however, preferable that there not be a significant amount of labeled homogeneous contaminating extraneous sequences. This constraint does usually require the purification of the target molecule to be sequenced, but a specific label technique would dispense with a purification requirement if the unlabeled extraneous sequences do not interfere with the labeled sequences.

In addition, conformational effects of target polypeptide folding may, in certain embodiments, be negligible if the polypeptide is fragmented into sufficiently small peptides, or if the interaction is performed under conditions where conformation, but not specific interaction, is disrupted.

B. Complications

Two obvious complications exist with the method of sequence analysis by hybridization. The first results from a probe of inappropriate length while the second relates to internally repeated sequences.

The first obvious complication is a problem which arises from an inappropriate length of recognition sequence, which causes problems with the specificity of recognition. For example, if the recognized sequence is too short, every sequence which is utilized will be recognized by every probe sequence. This occurs, e.g., in a binary system where the probes are each of sequences which occur relatively frequently, e.g., a two character probe for the binary system. Each possible two character probe would be expected to appear $\frac{1}{4}$ of the time in every single two character position. Thus, the above sequence example would be recognized by each of the 00, 10, 01, and 11. Thus, the sequence information is virtually lost because the resolution is too low and each recognition reagent specifically binds at multiple sites on the target sequence.

The number of different probes which bind to a target depends on the relationship between the probe length and the target length. At the extreme of short probe length, the just mentioned problem exists of excessive redundancy and lack of resolution. The lack of stability in recognition will also be a problem with extremely short probes. At the extreme of long probe length, each entire probe sequence is on a different position of a substrate. However, a problem arises from the number of possible sequences, which goes up dramatically with the length of the sequence. Also, the specificity of recognition begins to decrease as the contribution to binding by any particular subunit may become sufficiently low that the system fails to distinguish the fidelity of recognition. Mismatched hybridization may be a problem with the polynucleotide sequencing applications, though the fingerprinting and mapping applications may not be so strict in their fidelity requirements. As indicated above, a thirty position binary sequence has over a million possible sequences, a number which starts to become unreasonably large in its required number of different sequences, even though the target length is still very short. Preparing a substrate with all sequence possibilities for a long target may be extremely difficult due to the many different oligomers which must be synthesized.

The above example illustrates how a long target sequence may be reconstructed with a reasonably small number of shorter subsequences. Since the present day resolution of the regions of the substrate having defined oligomer probes

US 6,576,424 B2

17

attached to the substrate approaches about 10 microns by 10 microns for resolvable regions, about 10^6 , or 1 million, positions can be placed on a one centimeter square substrate. However, high resolution systems may have particular disadvantages which may be outweighed using the lower density substrate matrix pattern. For this reason, a sufficiently large number of probe sequences can be utilized so that any given target sequence may be determined by hybridization to a relatively small number of probes.

A second complication relates to convergence of sequences to a single subsequence. This will occur when a particular subsequence is repeated in the target sequence. This problem can be addressed in at least two different ways. The first, and simpler way, is to separate the repeat sequences onto two different targets. Thus, each single target will not have the repeated sequence and can be analyzed to its end. This solution, however, complicates the analysis by requiring that some means for cutting at a site between the repeats can be located. Typically a careful sequencer would want to have two intermediate cut points so that the intermediate region can also be sequenced in both directions across each of the cut points. This problem is inherent in the hybridization method for sequencing but can be minimized by using a longer known probe sequence so that the frequency of probe repeats is decreased.

Knowing the sequence of flanking sequences of the repeat will simplify the use of polymerase chain reaction (PCR) or a similar technique to further definitively determine the sequence between sequence repeats. Probes can be made to hybridize to those known sequences adjacent the repeat sequences, thereby producing new target sequences for analysis. See, e.g., Innis et al. (eds.) (1990) *PCR Protocols: A Guide to Methods and Applications*, Academic Press; and methods for synthesis of oligonucleotide probes, see, e.g., Gait (1984) *Oligonucleotide Synthesis: A Practical Approach*, IRL Press, Oxford.

Other means for dealing with convergence problems include using particular longer probes, and using degeneracy reducing analogues, see, e.g., Macevitz, S. (1990) PCT publication number WO 90/04652, which is hereby incorporated herein by reference. By use of stretches of the degeneracy reducing analogues with other probes in particular combinations, the number of probes necessary to fully saturate the possible oligomer probes is decreased. For example, with a stretch of 12-mers having the central 4-mer of degenerate nucleotides, in combination with all of the possible 8-mers, the collection numbers twice the number of possible 8-mers, e.g. $65,536+65,536=131,072$, but the population provides screening equivalent to all possible 12-mers.

By way of further explanation, all possible oligonucleotide 8-mers may be depicted in the fashion:

N1-N2-N3-N4-N5-N6-N7-N8,

in which there are $4^8=65,536$ possible 8-mers. As described in Ser. No. 07/624,120, now abandoned, producing all possible 8-mers requires $4 \times 8=32$ chemical binary synthesis steps to produce the entire matrix pattern of 65,536 8-mer possibilities. By incorporating degeneracy reducing nucleotides, D's, which hybridize nonselectively to any corresponding complementary nucleotide, new oligonucleotide 12-mers can be made in the fashion:

N1-N2-N3-N4-D-D-D-D-N5-N6-N7-N8,

in which there are again, as above, only $4^8=65,536$ possible "12-mers", which in reality only have 8 different nucleotides.

However, it can be seen that each possible 12-mer probe could be represented by a group of the two 8-mer types. Moreover, repeats of less than 12 nucleotides would not

18

converge, or cause repeat problems in the analysis. Thus, instead of requiring a collection of probes corresponding to all 12-mers, or $4^{12}=16,777,216$ different 12-mers, the same information can be derived by making 2 sets of "8-mers" consisting of the typical 8-mer collection of $4^8=65,536$ and the "12-mer" set with the degeneracy reducing analogues, also requiring making $4^8=65,536$. The combination of the two sets, requires making $65,536+65,536=131,072$ different molecules, but giving the information of 16,777,216 molecules. Thus, incorporating the degeneracy reducing analogue decreases the number of molecules necessary to get 12-mer resolution by a factor of about 128-fold.

C. Non-polynucleotide Embodiments

The above example is directed towards a polynucleotide embodiment. This application is relatively easily achieved because the specific reagents will typically be complementary oligonucleotides, although in certain embodiments other specific reagents may be desired. For example, there may be circumstances where other than complementary base pairing will be utilized. The polynucleotide targets, will usually be single strand, but may be double or triple stranded in various applications. However, a triple stranded specific interaction might be sometimes desired, or a protein or other specific binding molecule may be utilized. For example, various promoter or DNA sequence specific binding proteins might be used, including, e.g., restriction enzyme binding domains, other binding domains, and antibodies. Thus, specific recognition reagents besides oligonucleotides may be utilized.

For other polymer targets, the specific reagents will often be polypeptides. These polypeptides may be protein binding domains from enzymes or other proteins which display specificity for binding. Usually an antibody molecule may be used, and monoclonal antibodies may be particularly desired. Classical methods may be applied for preparing antibodies, see, e.g., Harlow and Lane (1988) *Antibodies: A Laboratory Manual* Cold Spring Harbor Press, New York; and Goding (1986) *Monoclonal Antibodies: Principles and Practice* (2d Ed.) Academic Press, San Diego. Other suitable techniques for in vitro exposure of lymphocytes to the antigens or selection of libraries of antibody binding sites are described, e.g., in Huse et al. (1989) *Science* 246:1275-1281; and Ward et al. (1989) *Nature* 341:544-546, each of which is hereby incorporated herein by reference. Unusual antibody production methods are also described, e.g., in Hendricks et al. (1989) *BioTechnology* 7:1271-1274; and Hiatt et al. (1989) *Nature* 342:76-78, each of which is hereby incorporated herein by reference. Other molecules which may exhibit specific binding interaction may be useful for attachment to a VLSIPS substrate by various methods, including the caged biotin methods, see, e.g., Ser. No. 07/435,316, now abandoned, and Barrett et al. (1993) U.S. Pat. No. 5,252,743.

The antibody specific reagents should be particularly useful for the polypeptide, carbohydrate, and synthetic polymer applications. Individual specific reagents might be generated by an automated process to generate the number of reagents necessary to advantageously use the high density positional matrix pattern. In an alternative approach, a plurality of hybridoma cells may be screened for their ability to bind to a VLSIPS matrix possessing the desired sequences whose binding specificity is desired. Each cell might be individually grown up and its binding specificity determined by VLSIPS apparatus and technology. An alternative strategy would be to expose the same VLSIPS matrix to a polyclonal serum of high titer. By a successively large volume of serum and different animals, each region of the VLSIPS substrate would have attached to it a substantial

US 6,576,424 B2

19

number of antibody molecules with specificity of binding. The substrate, with non-covalently bound antibodies could be derivatized and the antibodies transferred to an adjacent second substrate in the matrix pattern in which the antibody molecules had attached to the first matrix. If the sensitivity of detection of binding interaction is sufficiently high, such a low efficiency transfer of antibody molecules may produce a sufficiently high signal to be useful for many purposes, including the sequencing applications.

In another embodiment, capillary forces may be used to transfer the selected reagents to a new matrix, to which the reagents would be positionally attached in the pattern of the recognized sequences. Or, the reagents could be transversely electrophoresed, magnetically transferred, or otherwise transported to a new substrate in their retained positional pattern.

III. POLYNUCLEOTIDE SEQUENCING

In principle, the making of a substrate having a positionally defined matrix pattern of all possible oligonucleotides of a given length involves a conceptually simple method of synthesizing each and every different possible oligonucleotide, and affixing them to a definable position. Oligonucleotide synthesis is presently mechanized and enabled by current technology, see, e.g., Ser. No. 07/362, 901, now abandoned; Pirrung et al. (1992) U.S. Pat. No. 5,143,854; and instruments supplied by Applied Biosystems, Foster City, Calif.

A. Preparation of Substrate Matrix

The production of the collection of specific oligonucleotides used in polynucleotide sequencing may be produced in at least two different ways. Present technology certainly allows production of ten nucleotide oligomers on a solid phase or other synthesizing system. See, e.g., instrumentation provided by Applied Biosystems, Foster City, Calif. Although a single oligonucleotide can be relatively easily made, a large collection of them would typically require a fairly large amount of time and investment. For example, there are $4^{10}=1,048,576$ possible ten nucleotide oligomers. Present technology allows making each and every one of them in a separate purified form though such might be costly and laborious.

Once the desired repertoire of possible oligomer sequences of a given length have been synthesized, this collection of reagents may be individually positionally attached to a substrate, thereby allowing a batchwise hybridization step. Present technology also would allow the possibility of attaching each and every one of these 10-mers to a separate specific position on a solid matrix. This attachment could be automated in any of a number of ways, particularly through the use of a caged biotin type linking. This would produce a matrix having each of different possible 10-mers.

A batchwise hybridization is much preferred because of its reproducibility and simplicity. An automated process of attaching various reagents to positionally defined sites on a substrate is provided in Pirrung et al. (1992) U.S. Pat. No. 5,143,854; Ser. No. 07/624,120, now abandoned; and Barrett et al. (1993) U.S. Pat. No. 5,252,743; each of which is hereby incorporated herein by reference.

Instead of separate synthesis of each oligonucleotide, these oligonucleotides are conveniently synthesized in parallel by sequential synthetic processes on a defined matrix pattern as provided in Pirrung et al. (1992) U.S. Pat. No. 5,143,854; and Ser. No. 07/624,120, now abandoned, which are incorporated herein by reference. Here, the oligonucleotides are synthesized stepwise on a substrate at positionally

20

separate and defined positions. Use of photosensitive blocking reagents allows for defined sequences of synthetic steps over the surface of a matrix pattern. By use of the binary masking strategy, the surface of the substrate can be positioned to generate a desired pattern of regions, each having a defined sequence oligonucleotide synthesized and immobilized thereto.

Although the prior art technology can be used to generate the desired repertoire of oligonucleotide probes, an efficient and cost effective means would be to use the VLSIPS technology described in Pirrung et al. (1992) U.S. Pat. No. 5,143,854 and Ser. No. 07/624,120, now abandoned. In this embodiment, the photosensitive reagents involved in the production of such a matrix are described below.

The regions for synthesis may be very small, usually less than about $100\text{ }\mu\text{m}\times 100\text{ }\mu\text{m}$, more usually less than about $50\text{ }\mu\text{m}\times 50\text{ }\mu\text{m}$. The photolithography technology allows synthetic regions of less than about $10\text{ }\mu\text{m}\times 10\text{ }\mu\text{m}$, about $3\text{ }\mu\text{m}\times 3\text{ }\mu\text{m}$, or less. The detection also may detect such sized regions, though larger areas are more easily and reliably measured.

At a size of about 30 microns by 30 microns, one million regions would take about 11 centimeters square or a single wafer of about 4 centimeters by 4 centimeters. Thus the present technology provides for making a single matrix of that size having all one million plus possible oligonucleotides. Region size is sufficiently small to correspond to densities of at least about 5 regions/cm², 20 regions/cm², 50 regions/cm², 100 regions/cm², and greater, including 300 regions/cm², 1000 regions/cm², 3K regions/cm², 10K regions/cm², 30K regions/cm², 100K regions/cm², 300K regions/cm² or more, even in excess of one million regions/cm².

Although the pattern of the regions which contain specific sequences is theoretically not important, for practical reasons certain patterns will be preferred in synthesizing the oligonucleotides. The application of binary masking algorithms for generating the pattern of known oligonucleotide probes is described in related Ser. No. 07/624,120, now abandoned, which was filed simultaneously with this application. By use of these binary masks, a highly efficient means is provided for producing the substrate with the desired matrix pattern of different sequences. Although the binary masking strategy allows for the synthesis of all lengths of polymers, the strategy may be easily modified to provide only polymers of a given length. This is achieved by omitting steps where a subunit is not attached.

The strategy for generating a specific pattern may take any of a number of different approaches. These approaches are well described in related application Ser. No. 07/624, 120, now abandoned, and include a number of binary masking approaches which will not be exhaustively discussed herein. However, the binary masking and binary synthesis approaches provide a maximum of diversity with a minimum number of actual synthetic steps.

The length of oligonucleotides used in sequencing applications will be selected on criteria determined to some extent by the practical limits discussed above. For example, if probes are made as oligonucleotides, there will be 65,536 possible eight nucleotide sequences. If a nine subunit oligonucleotide is selected, there are 262,144 possible permutations of sequences. If a ten-mer oligonucleotide is selected, there are 1,048,576 possible permutations of sequences. As the number gets larger, the required number of positionally defined subunits necessary to saturate the possibilities also increases. With respect to hybridization conditions, the length of the matching necessary to confer stability of the

US 6,576,424 B2

21

conditions selected can be compensated for. See, e.g., Kanehisa, M. (1984) *Nuc. Acids Res.* 12:203–213, which is hereby incorporated herein by reference.

Although not described in detail here, but below for oligonucleotide probes, the VLSIPS technology would typically use a photosensitive protective group on an oligonucleotide. Sample oligonucleotides are shown in FIG. 1. In particular, the photoprotective group on the nucleotide molecules may be selected from a wide variety of positive light reactive groups preferably including nitro aromatic compounds such as o-nitro-benzyl derivatives or benzylsulfonyl. See, e.g., Gait (1984) *Oligonucleotide Synthesis: A Practical Approach*, IRL Press, Oxford, which is hereby incorporated herein by reference. In a preferred embodiment, 6-nitro-veratryl oxycarbonyl (NVOC), 2-nitrobenzyl oxycarbonyl (NBOC), or α,α -dimethyl-dimethoxybenzyl oxycarbonyl (DEZ) is used. Photoremovable protective groups are described in, e.g., Patchornik (1970) *J. Amer. Chem. Soc.* 92:6333–6335; and Amit et al. (1974) *J. Organic Chem.* 39:192–196; each of which is hereby incorporated herein by reference.

A preferred linker for attaching the oligonucleotide to a silicon matrix is illustrated in FIG. 2. A more detailed description is provided below. A photosensitive blocked nucleotide may be attached to specific locations of unblocked prior cycles of attachments on the substrate and can be successively built up to the correct length oligonucleotide probe.

It should be noted that multiple substrates may be simultaneously exposed to a single target sequence where each substrate is a duplicate of one another or where, in combination, multiple substrates together provide the complete or desired subset of possible subsequences. This provides the opportunity to overcome a limitation of the density of positions on a single substrate by using multiple substrates. In the extreme case, each probe might be attached to a single bead or substrate and the beads sorted by whether there is a binding interaction. Those beads which do bind might be encoded to indicate the subsequence specificity of reagents attached thereto.

Then, the target may be bound to the whole collection of beads and those beads that have appropriate specific reagents on them will bind to the target. Then a sorting system may be utilized to sort those beads that actually bind the target from those that do not. This may be accomplished by presently available cell sorting devices or a similar apparatus. After the relatively small number of beads which have bound the target have been collected, the encoding scheme may be read off to determine the specificity of the reagent on the bead. An encoding system may include a magnetic system, a shape encoding system, a color encoding system, or a combination of any of these, or any other encoding system. Once again, with the collection of specific interactions that have occurred, the binding may be analyzed for sequence information, fingerprint information, or mapping information.

The parameters of polynucleotide sizes of both the probes and target sequences are determined by the applications and other circumstances. The length of the oligonucleotide probes used will depend in part upon the limitations of the VLSIPS technology to provide the number of desired probes. For example, in an absolute sequencing application, it is often useful to have virtually all of the possible oligonucleotides of a given length. As indicated above, there are 65,536 8-mers, 262,144 9-mers, 1,048,576 10-mers, 4,194,304 11-mers, etc. As the length of the oligomer increases the number of different probes which must be

22

synthesized also increases at a rate of a factor of 4 for every additional nucleotide. Eventually the size of the matrix and the limitations in the resolution of regions in the matrix will reach the point where an increase in number of probes becomes disadvantageous. However, this sequencing procedure requires that the system be able to distinguish, by appropriate selection of hybridization and washing conditions, between binding of absolute fidelity and binding of complementary sequences containing mismatches. On the other hand, if the fidelity is unnecessary, this discrimination is also unnecessary and a significantly longer probe may be used. Significantly longer probes would typically be useful in fingerprinting or mapping applications.

The length of the probe is selected for a length that will allow the probe to bind with specificity to possible targets. The hybridization conditions are also very important in that they will determine how closely the homology of complementary binding will be detected. In fact, a single target may be evaluated at a number of different conditions to determine its spectrum of specificity for binding particular probes. This may find use in a number of other applications besides the polynucleotide sequencing fingerprinting or mapping. For example, it will be desired to determine the spectrum of binding affinities and specificities of cell surface antigens with binding by particular antibodies immobilized on the substrate surface, particularly under different interaction conditions. In a related fashion, different regions with reagents having differing affinities or levels of specificity may allow such a spectrum to be defined using a single incubation, where various regions, at a given hybridization condition, show the binding affinity. For example, fingerprint probes of various lengths, or with specific defined non-matches may be used. Unnatural nucleotides or nucleotides exhibiting modified specificity of complementary binding are described in greater detail in Macevicz (1990) PCT pub. No. WO 90/04652; and see the section on modified nucleotides in the Sigma Chemical Company catalogue. B. Labeling Target Nucleotide

The label used to detect the target sequences will be determined, in part, by the detection methods being applied. Thus, the labeling method and label used are selected in combination with the actual detecting systems being used.

Once a particular label has been selected, appropriate labeling protocols will be applied, as described below for specific embodiments. Standard labeling protocols for nucleic acids are described, e.g., in Sambrook et al.; Kambara, H. et al. (1988) *BioTechnology* 6:816–821; Smith, L. et al. (1985) *Nuc. Acids Res.* 13:2399–2412; for polypeptides, see, e.g., Allen G. (1989) *Sequencing of Proteins and Peptides*, Elsevier, N.Y., especially chapter 5, and Greenstein and Winitz (1961) *Chemistry of the Amino Acids*, Wiley and Sons, New York. Carbohydrate labeling is described, e.g., in Chaplin and Kennedy (1986) *Carbohydrate Analysis: A Practical Approach*, IRL Press, Oxford. Labeling of other polymers will be performed by methods applicable to them as recognized by a person having ordinary skill in manipulating the corresponding polymer.

In some embodiments, the target need not actually be labeled if a means for detecting where interaction takes place is available. As described below, for a nucleic acid embodiment, such may be provided by an intercalating dye which intercalates only into double stranded segments, e.g., where interaction occurs. See, e.g., Sheldon et al. U.S. Pat. No. 4,582,789.

In many uses, the target sequence will be absolutely homogeneous, both with respect to the total sequence and with respect to the ends of each molecule. Homogeneity

US 6,576,424 B2

23

with respect to sequence is important to avoid ambiguity. It is preferable that the target sequences of interest not be contaminated with a significant amount of labeled contaminating sequences. The extent of allowable contamination will depend on the sensitivity of the detection system and the inherent signal to noise of the system. Homogeneous contamination sequences will be particularly disruptive of the sequencing procedure.

However, although the target polynucleotide must have a unique sequence, the target molecules need not have identical ends. In fact, the homogeneous target molecule preparation may be randomly sheared to increase the numerical number of molecules. Since the total information content remains the same, the shearing results only in a higher number of distinct sequences which may be labeled and bind to the probe. This fragmentation may give a vastly superior signal relative to a preparation of the target molecules having homogeneous ends. The signal for the hybridization is likely to be dependent on the numerical frequency of the target-probe interactions. If a sequence is individually found on a larger number of separate molecules a better signal will result. In fact, shearing a homogeneous preparation of the target may often be preferred before the labeling procedure is performed, thereby producing a large number of labeling groups associated with each subsequence.

C. Hybridization Conditions

The hybridization conditions between probe and target should be selected such that the specific recognition interaction, i.e., hybridization, of the two molecules is both sufficiently specific and sufficiently stable. See, e.g., Hames and Higgins (1985) *Nucleic Acid Hybridisation: A Practical Approach*, IRL Press, Oxford. These conditions will be dependent both on the specific sequence and often on the guanine and cytosine (GC) content of the complementary hybrid strands. The conditions may often be selected to be universally equally stable independent of the specific sequences involved. This typically will make use of a reagent such as an alkylammonium buffer. See, Wood et al. (1985) "Base Composition-independent Hybridization in Tetramethylammonium Chloride: A Method for Oligonucleotide Screening of Highly Complex Gene Libraries," *Proc. Natl. Acad. Sci. USA*, 82:1585-1588; and Krupov et al. (1989) "An Oligonucleotide Hybridization Approach to DNA Sequencing," *FEBS Letters*, 256:118-122; each of which is hereby incorporated herein by reference. An alkylammonium buffer tends to minimize differences in hybridization rate and stability due to GC content. By virtue of the fact that sequences then hybridize with approximately equal affinity and stability, there is relatively little bias in strength or kinetics of binding for particular sequences. Temperature and salt conditions along with other buffer parameters should be selected such that the kinetics of renaturation should be essentially independent of the specific target subsequence or oligonucleotide probe involved. In order to ensure this, the hybridization reactions will usually be performed in a single incubation of all the substrate matrices together exposed to the identical same target probe solution under the same conditions.

Alternatively, various substrates may be individually treated differently. Different substrates may be produced, each having reagents which bind to target subsequences with substantially identical stabilities and kinetics of hybridization. For example, all of the high GC content probes could be synthesized on a single substrate which is treated accordingly. In this embodiment, the alkylammonium buffers could be unnecessary. Each substrate is then treated in a manner such that the collection of substrates show essentially uni-

24

form binding and the hybridization data of target binding to the individual substrate matrix is combined with the data from other substrates to derive the necessary subsequence binding information. The hybridization conditions will usually be selected to be sufficiently specific such that the fidelity of base matching will be properly discriminated. Of course, control hybridizations should be included to determine the stringency and kinetics of hybridization.

D. Detection; VLSIPS™ Technology Scanning

The next step of the sequencing process by hybridization involves labeling of target polynucleotide molecules. A quickly and easily detectable signal is preferred. The VLSIPS™ Technology apparatus is designed to easily detect a fluorescent label, so fluorescent tagging of the target sequence is preferred. Other suitable labels include heavy metal labels, magnetic probes, chromogenic labels (e.g., phosphorescent labels, dyes, and fluorophores) spectroscopic labels, enzyme linked labels, radioactive labels, and labeled binding proteins. Additional labels are described in U.S. Pat. No. 4,366,241, which is incorporated herein by reference.

The detection methods used to determine where hybridization has taken place will typically depend upon the label selected above. Thus, for a fluorescent label a fluorescent detection step will typically be used. Pirrung et al. (1992) U.S. Pat. No. 5,143,854 and Ser. No. 07/624,120, now abandoned, describe apparatus and mechanisms for scanning a substrate matrix using fluorescence detection, but a similar apparatus is adaptable for other optically detectable labels.

The detection method provides a positional localization of the region where hybridization has taken place. However, the position is correlated with the specific sequence of the probe since the probe has specifically been attached or synthesized at a defined substrate matrix position. Having collected all of the data indicating the subsequences present in the target sequence, this data may be aligned by overlap to reconstruct the entire sequence of the target, as illustrated above.

It is also possible to dispense with actual labeling if some means for detecting the positions of interaction between the sequence specific reagent and the target molecule are available. This may take the form of an additional reagent which can indicate the sites either of interaction, or the sites of lack of interaction, e.g., a negative label. For the nucleic acid embodiments, locations of double strand interaction may be detected by the incorporation of intercalating dyes, or other reagents such as antibody or other reagents that recognize helix formation, see, e.g., Sheldon, et al. (1986) U.S. Pat. No. 4,582,789, which is hereby incorporated herein by reference.

E. Analysis

Although the reconstruction can be performed manually as illustrated above, a computer program will typically be used to perform the overlap analysis. A program may be written and run on any of a large number of different computer hardware systems. The variety of operating systems and languages useable will be recognized by a computer software engineer. Various different languages may be used, e.g., BASIC; C; PASCAL; etc. A simple flow chart of data analysis is illustrated in FIG. 1.

F. Substrate Reuse

Finally, after a particular sequence has been hybridized and the pattern of hybridization analyzed, the matrix substrate should be reusable and readily prepared for exposure to a second or subsequent target polynucleotides. In order to do so, the hybrid duplexes are disrupted and the matrix

US 6,576,424 B2

25

treated in a way which removes all traces of the original target. The matrix may be treated with various detergents or solvents to which the substrate, the oligonucleotide probes, and the linkages to the substrate are inert. This treatment may include an elevated temperature treatment, treatment with organic or inorganic solvents, modifications in pH, and other means for disrupting specific interaction. Thereafter, a second target may actually be applied to the recycled matrix and analyzed as before.

G. Non-Polynucleotide Aspects

Although the sequencing, fingerprinting, and mapping functions will make use of the natural sequence recognition property of complementary nucleotide sequences, the non-polynucleotide sequences typically require other sequence recognition reagents. These reagents will take the form, typically, of proteins exhibiting binding specificity, e.g., enzyme binding sites or antibody binding sites.

Enzyme binding sites may be derived from promoter proteins, restriction enzymes, and the like. See, e.g., Stryer, L. (1988) *Biochemistry*, W.H. Freeman, Palo Alto. Antibodies will typically be produced using standard procedures, see, e.g., Harlow and Lane (1988) *Antibodies: A Laboratory Manual*, Cold Spring Harbor Press, New York; and Goding (1986) *Monoclonal Antibodies: Principles and Practice*, (2d Ed.) Academic Press, San Diego.

Typically, an antigen, or collection of antigens are presented to an immune system. This may take the form of synthesized short polymers produced by the VLSIPS technology, or by the other synthetic means, or from isolation of natural products. For example, antigen for the polypeptides may be made by the VLSIPS technology, by standard peptide synthesis,

The antigen or collection is presented to an appropriate immune system, e.g., to a whole animal as in a standard immunization protocol, or to a collection of immune cells or equivalent. In particular, see Ward et al. (1989) *Nature* 341:544-546; and Huse et al. (1989) *Science* 246:1275-1281, each of which is hereby incorporated herein by reference.

A large diversity of antibodies will be generated, some of which have specificities for the desired sequences. Antibodies may be purified having the desired sequence specificities by isolating the cells producing them. For example, a VLSIPS substrate with the desired antigens synthesized thereon may be used to isolate cells with cell surface reagents which recognize the antigens. The VLSIPS substrate may be used as an affinity reagent to select and recover the appropriate cells. Antibodies from those cells may be attached to a substrate using the caged biotin methodology, or by attaching a targeting molecule, e.g., an oligonucleotide. Alternatively, the supernatants from antibody producing cells can be easily assayed using a VLSIPS substrate to identify the cells producing the appropriate antibodies.

Although cells may be isolated, specific antibody molecules which perform the sequence recognition will also be sufficient. Preferably populations of antibody with a known specificity can be isolated. Supernatants from a large population of producing cells may be passed over a VLSIPS substrate to bind to the desired antigens attached to the substrate. When a sufficient density of antibody molecules are attached, they may be removed by an automated process, preferably as antibody populations exhibiting specificity of binding.

In one particular embodiment, a VLSIPS substrate, e.g., with a large plurality of fingerprint antigens attached thereto, is used to isolate antibodies from a supernatant of a population of cells producing antibodies to the antigens. Using

26

the substrate as an affinity reagent, the antibodies will attach to the appropriate positionally defined antigens. The antibodies may be carefully removed therefrom, preferably by an automated system which retains their homogeneous specificities. The isolated antibodies can be attached to a new substrate in a positionally defined matrix pattern.

In a further embodiment, these spatially separated antibodies may be isolated using a specific targeting method for isolation. In this embodiment, a linker molecule which attaches to a particular portion of the antibody, preferably away from the binding site, can be attached to the antibodies. Various reagents will be used, including staphylococcus protein A or antibodies which bind to domains remote from the binding site. Alternatively, the antibodies in the population, before affinity purification, may be derivatized with an appropriate reagent compatible with new VLSIPS synthesis. A preferred reagent is a nucleotide which can serve as a linker to synthetic VLSIPS steps for synthesizing a specific sequence thereon. Then, by successive VLSIPS cycles, each of the antibodies attached to the defined antigen regions can have a defined oligonucleotide synthesized thereon and corresponding in area to the region of the substrate having each antigen attached. These defined oligonucleotides will be useful as targeting reagents to attach those antibodies possessing the same target sequence specificity at defined positions on a new substrate, by virtue of having bound to the antigen region, to a new VLSIPS substrate having the complementary target oligonucleotides positionally located on it. In this fashion, a VLSIPS substrate having the desired antigens attached thereto can be used to generate a second VLSIPS substrate with positionally defined reagents which recognize those antigens.

The selected antigens will typically be selected to be those which define particular functionalities or properties, so as to be useful for fingerprinting and other uses. They will also be useful for mapping and sequencing embodiments.

IV. FINGERPRINTING

A. General

Many of the procedures and techniques used in the polynucleotide sequencing section are also appropriate for fingerprinting applications. See, e.g., Poustka, et al. (1986) *Cold Spring Harbor Symposia on Quant. Biol.*, vol. LI, 131-139, Cold Spring Harbor Press, New York; which is hereby incorporated herein by reference. The fingerprinting method provided herein is based, in part, upon the ability to positionally localize a large number of different specific probes onto a single substrate. This high density matrix pattern provides the ability to screen for, or detect, a very large number of different sequences simultaneously. In fact, depending upon the hybridization conditions, fingerprinting to the resolution of virtually absolute matching of sequence is possible thereby approaching an absolute sequencing embodiment. And the sequencing embodiment is very useful in identifying the probes useful in further fingerprinting uses. For example, characteristic features of genetic sequences will be identified as being diagnostic of the entire sequence. However, in most embodiments, longer probe and target will be used, and for which slight mismatching may not need to be resolved.

B. Preparation of Substrate Matrix

A collection of specific probes may be produced by either of the methods described above in the section on sequencing. Specific oligonucleotide probes of desired lengths may be individually synthesized on a standard oligonucleotide synthesizer. The length of these probes is limited only by the ability of the synthesizer to continue to accurately synthesize a molecule. Oligonucleotides or sequence fragments may

US 6,576,424 B2

27

also be isolated from natural sources. Biological amplification methods may be coupled with synthetic synthesizing procedures such as, e.g., polymerase chain reaction.

In one embodiment, the individually isolated probes may be attached to the matrix at defined positions. These probe reagents may be attached by an automated process making use of the caged biotin methodology described in Ser. No. 07/612,671, or using photochemical reagents, see, e.g., Dattagupta et al. (1985) U.S. Pat. No. 4,542,102 and (1987) U.S. Pat. No. 4,713,326. Each individually purified reagent can be attached individually at specific locations on a substrate.

In another embodiment, the VLSIPS synthesizing technique may be used to synthesize the desired probes at specific positions on a substrate. The probes may be synthesized by successively adding appropriate monomer subunits, e.g., nucleotides, to generate the desired sequences.

In another embodiment, a relatively short specific oligonucleotide is used which serves as a targeting reagent for positionally directing the sequence recognition reagent. For example, the sequence specific reagents having a separate additional sequence recognition segment (usually of a different polymer from the target sequence) can be directed to target oligonucleotides attached to the substrate. By use of non-natural targeting reagents, e.g., unusual nucleotide analogues which pair with other unnatural nucleotide analogues and which do not interfere with natural nucleotide interactions, the natural and non-natural portions can coexist on the same molecule without interfering with their individual functionalities. This can combine both a synthetic and biological production system analogous to the technique for targeting monoclonal antibodies to locations on a VLSIPS substrate at defined positions. Unnatural optical isomers of nucleotides may be useful unnatural reagents subject to similar chemistry, but incapable of interfering with the natural biological polymers. See also, Ser. No. 07/626,730, which is hereby incorporated herein by reference.

After the separate substrate attached reagents are attached to the targeting segment, the two are crosslinked, thereby permanently attaching them to the substrate. Suitable crosslinking reagents are known, see, e.g., Dattagupta et al. (1985) U.S. Pat. No. 4,542,102 and (1987) "Coupling of nucleic acids to solid support by photochemical methods," U.S. Pat. No. 4,713,326, each of which is hereby incorporated herein by reference. Similar linkages for attachment of proteins to a solid substrate are provided, e.g., in Merrifield (1986) *Science* 232:341-347, which is hereby incorporated herein by reference.

C. Labeling Target Nucleotides

The labeling procedures used in the sequencing embodiments will also be applicable in the fingerprinting embodiments. However, since the fingerprinting embodiments often will involve relatively large target molecules and relatively short oligonucleotide probes, the amount of signal necessary to incorporate into the target sequence may be less critical than in the sequencing applications. For example, a relatively long target with a relatively small number of labels per molecule may be easily amplified or detected because of the relatively large target molecule size.

In various embodiments, it may be desired to cleave the target into smaller segments as in the sequencing embodiments. The labeling procedures and cleavage techniques described in the sequencing embodiments would usually also be applicable here.

D. Hybridization Conditions

The hybridization conditions used in fingerprinting embodiments will typically be less critical than for the

28

sequencing embodiments. The reason is that the amount of mismatching which may be useful in providing the fingerprinting information would typically be far greater than that necessary in sequencing uses. For example, Southern hybridizations do not typically distinguish between slightly mismatched sequences. Under these circumstances, important and valuable information may be arrived at with less stringent hybridization conditions while providing valuable fingerprinting information. However, since the entire substrate is typically exposed to the target molecule at one time, the binding affinity of the probes should usually be of approximately comparable levels. For this reason, if oligonucleotide probes are being used, their lengths should be approximately comparable and will be selected to hybridize under conditions which are common for most of the probes on the substrate. Much as in a Southern hybridization, the target and oligonucleotide probes are of lengths typically greater than about 25 nucleotides. Under appropriate hybridization conditions, e.g., typically higher salt and lower temperature, the probes will hybridize irrespective of imperfect complementarity. In fact, with probes of greater than, e.g., about fifty nucleotides, the difference in stability of different sized probes will be relatively minor.

Typically the fingerprinting is merely for probing similarity or homology. Thus, the stringency of hybridization can usually be decreased to fairly low levels. See, e.g., Wetmur and Davidson (1968) "Kinetics of Renaturation of DNA," *J. Mol. Biol.*, 31:349-370; and Kanehisa, M. (1984) *Nuc. Acids Res.*, 12:203-213.

E. Detection; VLSIPS™ Technology Scanning

Detection methods will be selected which are appropriate for the selected label. The scanning device need not necessarily be digitized or placed into a specific digital database, though such would most likely be done. For example, the analysis in fingerprinting could be photographic. Where a standardized fingerprint substrate matrix is used, the pattern of hybridizations may be spatially unique and may be compared photographically. In this manner, each sample may have a characteristic pattern of interactions and the likelihood of identical patterns will preferably be such low frequency that the fingerprint pattern indeed becomes a characteristic pattern virtually as unique as an individual's fingertip fingerprint. With a standardized substrate, every individual could be, in theory, uniquely identifiable on the basis of the pattern of hybridizing to the substrate.

Of course, the VLSIPS™ Technology scanning apparatus may also be useful to generate a digitized version of the fingerprint pattern. In this way, the identification pattern can be provided in a linear string of digits. This sequence could also be used for a standardized identification system providing significant useful medical transferability of specific data. In one embodiment, the probes used are selected to be of sufficiently high resolution to measure the antigens of the major histocompatibility complex. It might even be possible to provide transplantation matching data in a linear stream of data. The fingerprinting data may provide a condensed version, or summary, of the linear genetic data, or any other information data base.

F. Analysis

The analysis of the fingerprint will often be much simpler than a total sequence determination. However, there may be particular types of analysis which will be substantially simplified by a selected group of probes. For example, probes which exhibit particular populational heterogeneity may be selected. In this way, analysis may be simplified and practical utility enhanced merely by careful selection of the specific probes and a careful matrix layout of those probes.

US 6,576,424 B2

29

G. Substrate Reuse

As with the sequencing application, the fingerprinting usages may also take advantage of the reusability of the substrate. In this way, the interactions can be disrupted, the substrate treated, and the renewed substrate is equivalent to an unused substrate.

H. Non-polynucleotide Aspects

Besides polynucleotide applications, the fingerprinting analysis may be applied to other polymers, especially polypeptides, carbohydrates, and other polymers, both organic and inorganic. Besides using the fingerprinting method for analyzing a particular polymer, the fingerprinting method may be used to characterize various samples. For example, a cell or population of cells may be tested for their expression of specific antigens or their mRNA sequence intent. For example, a T-cell may be classified by virtue of its combination of expressed surface antigens. With specific reagents which interact with these antigens, a cell or a population of cells or a lysed cell may be exposed to a VLSIPS substrate. The biological sample may be classified or characterized by analyzing the pattern of specific interaction. This may be applicable to a cell or tissue type, to the messenger RNA population expressed by a cell to the genetic content of a cell, or to virtually any sample which can be classified and/or identified by its combination of specific molecular properties.

The ability to generate a high density means for screening the presence or absence of specific interactions allows for the possibility of screening for, if not saturating, all of a very large number of possible interactions. This is very powerful in providing the means for testing the combinations of molecular properties which can define a class of samples. For example, a species of organism may be characterized by its DNA sequences, e.g., a genetic fingerprint. By using a fingerprinting method, it may be determined that all members of that species are sufficiently similar in specific sequences that they can be easily identified as being within a particular group. Thus, newly defined classes may be resolved by their similarity in fingerprint patterns. Alternatively, a non-member of that group will fail to share those many identifying characteristics. However, since the technology allows testing of a very large number of specific interactions, it also provides the ability to more finely distinguish between closely related different cells or samples. This will have important applications in diagnosing viral, bacterial, and other pathological or nonpathological infections.

In particular, cell classification may be defined by any of a number of different properties. For example, a cell class may be defined by its DNA sequences contained therein. This allows species identification for parasitic or other infections. For example, the human cell is presumably genetically distinguishable from a monkey cell, but different human cells will share many genetic markers. At higher resolution, each individual human genome will exhibit unique sequences that can define it as a single individual.

Likewise, a developmental stage of a cell type may be definable by its pattern of expression of messenger RNA. For example, in particular stages of cells, high levels of ribosomal RNA are found whereas relatively low levels of other types of messenger RNAs may be found. The high resolution distinguishability provided by this fingerprinting method allows the distinction between cells which have relatively minor differences in its expressed mRNA population. Where a pattern is shown to be characteristic of a stage, a stage may be defined by that particular pattern of messenger RNA expression.

30

In a similar manner, the antigenic determinants found on a protein may very well define the cell class. For example, immunological T-cells are distinguishable from B-cells because, in part, the cell surface antigens on the cell types are distinguishable. Different T-cell subclasses can be also distinguished from one another by whether they contain particular T-cell antigens. The present invention provides the possibility for high resolution testing of many different interactions simultaneously, and the definition of new cell types will be possible.

The high resolution VLSIPS™ substrate may also be used as a very powerful diagnostic tool to test the combination of presence, of a plurality of different assays from a biological sample. For example, a cancerous condition may be indicated by a combination of various different properties found in the blood. For example, a cancerous condition may be indicated by a combination of expression of various soluble antigens found in the blood along with a high number of various cellular antigens found on lymphocytes and/or particular cell degradation products. With a substrate as provided herein, a large number of different features can be simultaneously performed on a biological sample. In fact, the high resolution of the test will allow more complete characterization of parameters which define particular diseases. Thus, the power of diagnostic tests may be limited by the extent of statistical correlation with a particular condition rather than with the number of antigens or interactions which are tested. The present invention provides the means to generate this large universe of possible reagents and the ability to actually accumulate that correlative data.

In another embodiment, a substrate as provided herein may be used for genetic screening. This would allow for simultaneous screening of thousands of genetic markers. As the density of the matrix is increased, many more molecules can be simultaneously tested. Genetic screening then becomes a simpler method as the present invention provides the ability to screen for thousands, tens of thousands, and hundreds of thousands, even millions of different possible genetic features. However, the number of high correlation genetic markers for conditions numbers only in the hundreds. Again, the possibility for screening a large number of sequences provides the opportunity for generating the data which can provide correlation between sequences and specific conditions or susceptibility. The present invention provides the means to generate extremely valuable correlations useful for the genetic detection of the causative mutation leading to medical conditions. In still another embodiment, the present invention would be applicable to distinguishing two individuals having identical genetic compositions. The antibody population within an individual is dependent both on genetic and historical factors. Each individual experiences a unique exposure to various infectious agents, and the combined antibody expression is partly determined thereby. Thus, individuals may also be fingerprinted by their immunological content, either of actively expressed antibodies, or their immunological memory. Similar sorts of immunological and environmental histories may be useful for fingerprinting, perhaps in combination with other screening properties. In particular, the present invention may be useful for screening allergic reactions or susceptibilities, and a simple IgE specificity test may be useful in determining a spectrum of allergies.

With the definition of new classes of cells, a cell sorter will be used to purify them. Moreover, new markers for defining that class of cells will be identified. For example, where the class is defined by its RNA content, cells may be screened by antisense probes which detect the presence or

US 6,576,424 B2

31

absence of specific sequences therein. Alternatively, cell lysates may provide information useful in correlating intracellular properties with extracellular markers which indicate functional differences. Using standard cell sorter technology with a fluorescence or labeled antisense probe which recognizes the internal presence of the specific sequences of interest, the cell sorter will be able to isolate a relatively homogeneous population of cells possessing the particular marker. Using successive probes the sorting process should be able to select for cells having a combination of a large number of different markers.

In a non-polynucleotide embodiment, cells may be defined by the presence of other markers. The markers may be carbohydrates, proteins, or other molecules. Thus, a substrate having particular specific reagents, e.g., antibodies, attached to it should be able to identify cells having particular patterns of marker expression. Of course, combinations of these made be utilized and a cell class may be defined by a combination of its expressed mRNA, its carbohydrate expression, its antigens, and other properties. This fingerprinting should be useful in determining the physiological state of a cell or population of cells.

Having defined a cell type whose function or properties are defined by the reagents attachable to a VLSIPS substrate, such as cellular antigens, these structural manifestations of function may be used to sort cells to generate a relatively homogeneous population of that class of cells. Standard cell sorter technology may be applied to purify such a population, see, e.g., Dangel, J. and Herzenberg (1982) "Selection of hybridomas and hybridoma variants using the fluorescence activated cell sorter," *J. Immunological Methods* 52:1-14; and Becton Dickinson, Fluorescence Activated Cell Sorter Division, San Jose, Calif., and Coulter Diagnostics, Hialeah, Fla.

With the fingerprinting method an identification means arises from mosaicism problems in an organism. A mosaic organism is one whose genetic content in different cells is significantly different. Various clonal populations should have similar genetic fingerprints, though different clonal populations may have different genetic contents. See, for example, Suzuki et al. *An Introduction to Genetic Analysis* (4th Ed.), Freeman and Co., New York, which is hereby incorporated herein by reference. However, this problem should be a relatively rare problem and could be more carefully evaluated with greater experience using the fingerprinting methods.

The invention will also find use in detecting changes, both genetic and antigenic, e.g., in a rapidly "evolving" protozoa infection, or similarly changing organism.

V. MAPPING

A. General

The use of the present invention for mapping parallels its use for fingerprinting and sequencing. Where a polymer is a linear molecule, the mapping provides the ability to locate particular segments along the length of the polymer. Branched polymers can be treated as a series of individual linear polymers. The mapping provides the ability to locate, in a relative sense, the order of various subsequences. This may be achieved using at least two different approaches.

The first approach is to take the large sequence and fragment it at specific points. The fragments are then ordered and attached to a solid substrate. For example, the clones resulting from a chromosome walking process may be individually attached to the substrate by methods, e.g., caged biotin techniques, indicated earlier. Segments of unknown map position will be exposed to the substrate and will hybridize to the segment which contains that particular

32

sequence. This procedure allows the rapid determination of a number of different labeled segments, each mapping requiring only a single hybridization step once the substrate is generated. The substrate may be regenerated by removal of the interaction, and the next mapping segment applied.

In an alternative method, a plurality of subsequences can be attached to a substrate. Various short probes may be applied to determine which segments may contain particular overlaps. The theoretical basis and a description of this mapping procedure is contained in, e.g., Evans et al. 1989 "Physical Mapping of Complex Genomes by Cosmid Multiplex Analysis," *Proc. Natl. Acad. Sci. USA* 86:5030-5034, and other references cited above in the Section labeled "Overall Description." Using this approach, the details of the mapping embodiment are very similar to those used in the fingerprinting embodiment.

B. Preparation of Substrate Matrix

The substrate may be generated in either of the methods generally applicable in the sequencing and fingerprinting embodiments. The substrate may be made either synthetically, or by attaching otherwise purified probes or sequences to the matrix. The probes or sequences may be derived either from synthetic or biological means. As indicated above, the solid phase substrate synthetic methods may be utilized to generate a matrix with positionally defined sequences. In the mapping embodiment, the importance of saturation of all possible subsequences of a preselected length is far less important than in the sequencing embodiment, but the length of the probes used may be desired to be much longer. The processes for making a substrate which has longer oligonucleotide probes should not be significantly different from those described for the sequencing embodiments, but the optimization parameters may be modified to comply with the mapping needs.

C. Labeling

The labeling methods will be similar to those applicable in sequencing and fingerprinting embodiments. Again, it may be desirable to fragment the target sequences.

D. Hybridization/Specific Interaction

The specificity of interaction between the targets and probe would typically be closer to those used for fingerprinting embodiments, where homology is more important than absolute distinguishability of high fidelity complementary hybridization. Usually, the hybridization conditions will be such that merely homologous segments will interact and provide a positive signal. Much like the fingerprinting embodiment, it may be useful to measure the extent of homology by successive incubations at higher stringency conditions. Or, a plurality of different probes, each having various levels of homology may be used. In either way, the spectrum of homologies can be measured.

Where non-nucleic acid hybridization is involved, the specific interactions may also be compared in a fingerprint-like manner. The specific reagents may have less specificity, e.g., monoclonal antibodies which recognize a broader spectrum of sequences may be utilized relative to a sequencing embodiment. Again, the specificity of interaction may be measured under various conditions of increasing stringency to determine the spectrum of matching across the specific probes selected, or a number of different stringency reagents may be included to indicate the binding affinity.

E. Detection

The detection methods used in the mapping procedure will be virtually identical to those used in the fingerprinting embodiment. The detection methods will be selected in combination with the labeling methods.

US 6,576,424 B2

33

F. Analysis

The analysis of the data in a mapping embodiment will typically be somewhat different from that in fingerprinting. The fingerprinting embodiment will test for the presence or absence of specific or homologous segments. However, in the mapping embodiment, the existence of an interaction is coupled with some indication of the location of the interaction. The interaction is mapped in some manner to the physical polymer sequence. Some means for determining the relative positions of different probes is performed. This may be achieved by synthesis of the substrate in pattern, or may result from analysis of sequences after they have been attached to the substrate.

For example, the probes may be randomly positioned at various locations on the substrate. However, the relative positions of the various reagents in the original polymer may be determined by using short fragments, e.g., individually, as target molecules which determine the proximity of different probes. By an automated system of testing each different short fragment of the original polymer, coupled with proper analysis, it will be possible to determine which probes are adjacent one another on the original target sequence and correlate that with positions on the matrix. In this way, the matrix is useful for determining the relative locations of various new segments in the original target molecule. This sort of analysis is described in Evans, and the related references described above.

G. Substrate Reuse

The substrate should be reusable in the manner described in the fingerprinting section. The substrate is renewed by removal of the specific interactions and is washed and prepared for successive cycles of exposure to new target sequences.

H. Non-polynucleotide Aspects

The mapping procedure may be used on other molecules than polynucleotides. Although hybridization is one type of specific interaction which is clearly useful for use in this mapping embodiment, antibody reagents may also be very useful. In the same way that polypeptide sequencing or other polymers may be sequenced by the reagents and techniques described in the sequencing section and fingerprinting section, the mapping embodiment may also be used similarly.

In another form of mapping, as described above in the fingerprinting section, the developmental map of a cell or biological system may be measured using fingerprinting type technology. Thus, the mapping may be along a temporal dimension rather than along a polymer dimension. The mapping or fingerprinting embodiments may also be used in determining the genetic rearrangements which may be genetically important, as in lymphocyte and B-cell development. In another example, various rearrangements or chromosomal dislocations may be tested by either the fingerprinting or mapping methods. These techniques are similar in many respects and the fingerprinting and mapping embodiments may overlap in many respects.

VI. ADDITIONAL SCREENING AND APPLICATIONS

A. Specific Interactions

As originally indicated in the parent filing of VLSIPS™ Technology, the production of a high density plurality of spatially segregated polymers provides the ability to generate a very large universe or repertoire of individually and distinct sequence possibilities. As indicated above, particular oligonucleotides may be synthesized in automated fashion at specific locations on a matrix. In fact, these oligonucleotides may be used to direct other molecules to specific

34

locations by linking specific oligonucleotides to other reagents which are in batch exposed to the matrix and hybridized in a complementary fashion to only those locations where the complementary oligonucleotide has been synthesized on the matrix. This allows for spatially attaching a plurality of different reagents onto the matrix instead of individually attaching each separate reagent at each specific location. Although the caged biotin method allows automated attachment, the speed of the caged biotin attachment process is relatively slow and requires a separate reaction for each reagent being attached. By use of the oligonucleotide method, the specificity of position can be done in an automated and parallel fashion. As each reagent is produced, instead of directly attaching each reagent at each desired position, the reagent may be attached to a specific desired complementary oligonucleotide which will ultimately be specifically directed toward locations on the matrix having a complementary oligonucleotide attached thereat.

In addition, the technology allows screening for specificity of interaction with particular reagents. For example, the oligonucleotide sequence specificity of binding of a potential reagent may be tested by presenting to the reagent all of the possible subsequences available for binding. Although secondary or higher order sequence specific features might not be easily screenable using this technology, it does provide a convenient, simple, quick, and thorough screen of interactions between a reagent and its target recognition sequences. See, e.g., Pfeifer et al. (1989) *Science* 246:810-812.

For example, the interaction of a promoter protein with its target binding sequence may be tested for many different, or all, possible binding sequences. By testing the strength of interactions under various different conditions, the interaction of the promoter protein with each of the different potential binding sites may be analyzed. The spectrum of strength of interactions with each different potential binding site may provide significant insight into the types of features which are important in determining specificity.

An additional example of a sequence specific interaction between reagents is the testing of binding of a double stranded nucleic acid structure with a single stranded oligonucleotide. Often, a triple stranded structure is produced which has significant aspects of sequence specificity. Testing of such interactions with either sequences comprising only natural nucleotides, or perhaps the testing of nucleotide analogs may be very important in screening for particularly useful diagnostic or therapeutic reagents. See, e.g., Häner and Dervan (1990) *Biochemistry* 29:9761-6765, and references therein.

B. Sequence Comparisons

Once a gene is sequenced, the present invention provides a means to compare alleles or related sequences to locate and identify differences from the control sequence. This would be extremely useful in further analysis of genetic variability at a specific gene locus.

C. Categorizations

As indicated above in the fingerprinting and mapping embodiments, the present invention is also useful in defining specific stages in the temporal sequence of cells, e.g., development, and the resulting tissues within an organism. For example, the developmental stage of a cell, or population of cells, can be dependent upon the expression of particular messenger RNAs or cellular antigens. The screening procedures provided allow for high resolution definition of new classes of cells. In addition, the temporal development of particular cells will be characterized by the presence or expression of various mRNAs. Means to simultaneously

US 6,576,424 B2

35

screen a plurality or very large number of different sequences are provided. The combination of different markers made available dramatically increases the ability to distinguish fairly closely related cell types. Other markers may be combined with markers and methods made available herein to define new classifications of biological samples, e.g., based upon new combinations of markers.

The presence or absence of particular marker sequences will be used to define temporal developmental stages. Once the stages are defined, fairly simple methods can be applied to actually purify those particular cells. For example, antisense probes or recognition reagents may be used with a cell sorter to select those cells containing or expressing the critical markers. Alternatively, the expression of those sequences may result in specific antigens which may also be used in defining cell classes and sorting those cells away from others. In this way, for example, it should be possible to select a class of omnipotent immune system cells which are able to completely regenerate a human immune system. Based upon the cellular classes defined by the parameters made available by this technology, purified classes of cells having identifiable differences, structural or functional, are made available.

In an alternative embodiment, a plurality of antigens or specific binding proteins attached to the substrate may be used to define particular cell types. For example, subclasses of T-cells are defined, in part, by the combination of expressed cell surface antigens. The present invention allows for the simultaneous screening of a large plurality of different antigens together. Thus, higher resolution classification of different T-cell subclasses becomes possible and, with the definitions and functional differences which correlate with those antigenic or other parameters, the ability to purify those cell types becomes available. This is applicable not only to T-cells, but also to lymphocyte cells, or even to freely circulating cells. Many of the cells for which this would be most useful will be immobile cells found in particular tissues or organs. Tumor cells will be diagnosed or detected using these fingerprinting techniques. Coupled with a temporal change in structure, developmental classes may also be selected and defined using these technologies. The present invention also provides the ability not only to define new classes of cells based upon functional or structural differences, but it also provides the ability to select or purify populations of cells which share these particular properties. Standard cell sorting procedures using antibody markers may be used to detect extracellular features. Intracellular features would also be detectable by introducing the label reagents into the cell. In particular, antisense DNA or RNA molecules may be introduced into a cell to detect RNA sequences therein. See, e.g., Weintraub (1990) *Scientific American* 262:40-46.

D. Statistical Correlations

In an additional embodiment, the present invention also allows for the high resolution correlation of medical conditions with various different markers. For example, the presently available technology, when applied to amniocentesis or other genetic screening methods, typically screens for tens of different markers at most. The present invention allows simultaneous screening for tens, hundreds, thousands, tens of thousands, hundreds of thousands, and even millions of different genetic sequences. Thus, applying the fingerprinting methods of the present invention to a sufficiently large population allows detailed statistical analysis to be made, thereby correlating particular medical conditions with particular markers, typically antigenic or genetic. Tumor specific antigens will be identified using the present invention.

36

Various medical conditions may be correlated against an enormous data base of the sequences within an individual. Genetic propensities and correlations then become available and high resolution genetic predictability and correlation become much more easily performed. With the enormous data base, the reliability of the predictions is also better tested. Particular markers which are partially diagnostic of particular medical conditions or medical susceptibilities will be identified and provide direction in further studies and more careful analysis of the markers involved. Of course, as indicated above in the sequencing embodiment, the present invention will find much use in intense sequencing projects. For example, sequencing of the entire human genome in the human genome project will be greatly simplified and enabled by the present invention.

VI. FORMATION OF SUBSTRATE

The substrate is provided with a pattern of specific reagents which are positionally localized on the surface of the substrate. This matrix of positions is defined by the automated system which produces the substrate. The instrument will typically be one similar to that described in Pirrung et al. (1992) U.S. Pat. No. 5,143,854, and Ser. No. 07/624,120, now abandoned. The instrumentation described therein is directly applicable to the applications used here. In particular, the apparatus comprises a substrate, typically a silicon containing substrate, on which positions on the surface may be defined by a coordinate system of positions. These positions can be individually addressed or detected by the VLSIPS™ Technology apparatus.

Typically, the VLSIPS™ Technology apparatus uses optical methods used in semiconductor fabrication applications. In this way, masks may be used to photo-activate positions for attachment or synthesis of specific sequences on the substrate. These manipulations may be automated by the types of apparatus described in Pirrung et al. (1992) U.S. Pat. No. 5,143,854 and Ser. No. 07/624,120, now abandoned.

Selectively removable protecting groups allow creation of well defined areas of substrate surface having differing reactivities. Preferably, the protecting groups are selectively removed from the surface by applying a specific activator, such as electromagnetic radiation of a specific wavelength and intensity. More preferably, the specific activator exposes selected areas of surface to remove the protecting groups in the exposed areas.

Protecting groups of the present invention are used in conjunction with solid phase oligomer syntheses, such as peptide syntheses using natural or unnatural amino acids, nucleotide syntheses using deoxyribonucleic and ribonucleic acids, oligosaccharide syntheses, and the like. In addition to protecting the substrate surface from unwanted reaction, the protecting groups block a reactive end of the monomer to prevent self-polymerization. For instance, attachment of a protecting group to the amino terminus of an activated amino acid, such as the N-hydroxysuccinimide-activated ester of the amino acid prevents the amino terminus of one monomer from reacting with the activated ester portion of another during peptide synthesis.

Alternatively, the protecting group may be attached to the carboxyl group of an amino acid to prevent reaction at this site. Most protecting groups can be attached to either the amino or the carboxyl group of an amino acid, and the nature of the chemical synthesis will dictate which reactive group will require a protecting group. Analogously, attachment of a protecting group to the 5'-hydroxyl group of a nucleoside

US 6,576,424 B2

37

during synthesis using for example, phosphate-triester coupling chemistry, prevents the 5'-hydroxyl of one nucleoside from reacting with the 3'-activated phosphate-triester of another.

Regardless of the specific use, protecting groups are employed to protect a moiety on a molecule from reacting with another reagent. Protecting groups of the present invention have the following characteristics: they prevent selected reagents from modifying the group to which they are attached; they are stable (that is, they remain attached) to the synthesis reaction conditions; they are removable under conditions that do not adversely affect the remaining structure; and once removed, do not react appreciably with the surface or surface-bound oligomer. The selection of a suitable protecting group will depend, of course, on the chemical nature of the monomer unit and oligomer, as well as the specific reagents they are to protect against.

In a preferred embodiment, the protecting groups will be photoactivatable. The properties and uses of photoreactive protecting compounds have been reviewed. See, McCray et al., *Ann. Rev. of Biophys. and Biophys. Chem.* (1989) 18:239-270, which is incorporated herein by reference. Preferably, the photosensitive protecting groups will be removable by radiation in the ultraviolet (UV) or visible portion of the electromagnetic spectrum. More preferably, the protecting groups will be removable by radiation in the near UV or visible portion of the spectrum. In some embodiments, however, activation may be performed by other methods such as localized heating, electron beam lithography, laser pumping, oxidation or reduction with microelectrodes, and the like. Sulfonyl compounds are suitable reactive groups for electron beam lithography. Oxidative or reductive removal is accomplished by exposure of the protecting group to an electric current source, preferably using microelectrodes directed to the predefined regions of the surface which are desired for activation. A more detailed description of these protective groups is provided in Ser. No. 07/624,120, now abandoned, which is hereby incorporated herein by reference.

The density of reagents attached to a silicon substrate may be varied by standard procedures. The surface area for attachment of reagents may be increased by modifying the silicon surface. For example, a matte surface may be machined or etched on the substrate to provide more sites for attachment of the particular reagents. Another way to increase the density of reagent binding sites is to increase the derivitization density of the silicon. Standard procedures for achieving this are described, below.

One method to control the derivitization density is to highly derivatize the substrate with photochemical groups at high density. The substrate is then photolyzed for various predetermined times, which photoactivate the groups at a measurable rate, and react them with a capping reagent. By this method, the density of linker groups may be modulated by using a desired time and intensity of photoactivation.

In many applications, the number of different sequences which may be provided may be limited by the density and

38

the size of the substrate on which the matrix pattern is generated. In situations where the density is insufficiently high to allow the screening of the desired number of sequences, multiple substrates may be used to increase the number of sequences tested. Thus, the number of sequences tested may be increased by using a plurality of different substrates. Because the VLSIPS apparatus is almost fully automated, increasing the number of substrates does not lead to a significant increase in the number of manipulations which must be performed by humans. This again leads to greater reproducibility and speed in the handling of these multiple substrates.

A. Instrumentation

The concept of using VLSIPS™ Technology generally allows a pattern or a matrix of reagents to be generated. The procedure for making the pattern is performed by any of a number of different methods. An apparatus and instrumentation useful for generating a high density VLSIPS substrate is described in detail in Pirrung et al. (1992) U.S. Pat. No. 5,143,854 and Ser. No. 07/624,120, now abandoned.

B. Binary Masking

The details of the binary masking are described in an accompanying application filed simultaneously with this, Ser. No. 07/624,120, now abandoned, whose specification is incorporated herein by reference.

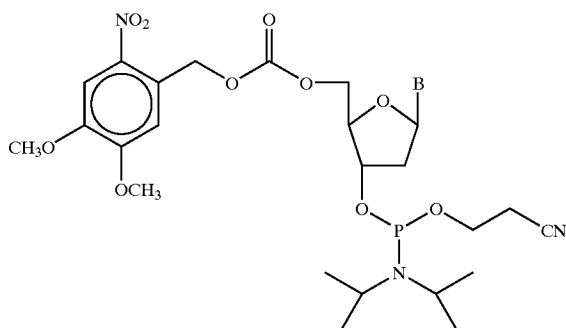
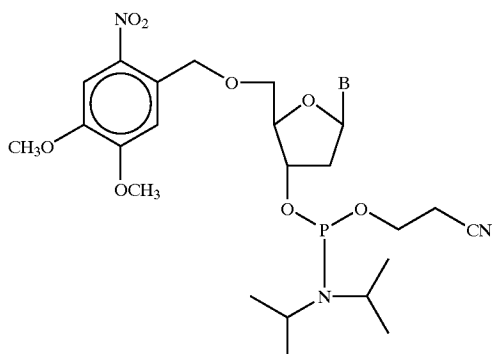
For example, the binary masking technique allows for producing a plurality of sequences based on the selection of either of two possibilities at any particular location. By a series of binary masking steps, the binary decision may be the determination, on a particular synthetic cycle, whether or not to add any particular one of the possible subunits. By treating various regions of the matrix pattern in parallel, the binary masking strategy provides the ability to carry out spatially addressable parallel synthesis.

C. Synthetic Methods

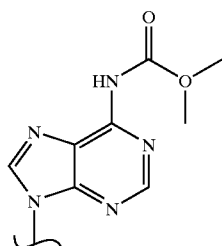
The synthetic methods in making a substrate are described in the parent application, Pirrung et al. (1992) U.S. Pat. No. 5,143,854. The construction of the matrix pattern on the substrate will typically be generated by the use of photo-sensitive reagents. By use of photo-lithographic optical methods, particular segments of the substrate can be irradiated with light to activate or deactivate blocking agents, e.g., to protect or deprotect particular chemical groups. By an appropriate sequence of photo-exposure steps at appropriate times with appropriate masks and with appropriate reagents, the substrates can have known polymers synthesized at positionally defined regions on the substrate. Methods for synthesizing various substrates are described in Pirrung et al. (1992) U.S. Pat. No. 5,143,854 and Ser. No. 07/624,120, now abandoned. By a sequential series of these photo-exposure and reaction manipulations, a defined matrix pattern of known sequences may be generated, and is typically referred to as a VLSIPS™ Technology substrate. In the nucleic acid synthesis embodiment, nucleosides used in the synthesis of DNA by photolytic methods will typically be one of the two forms shown below:

US 6,576,424 B2

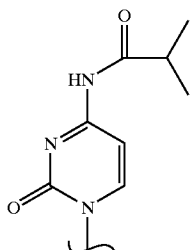
39



In I, the photolabile group at the 5' position is abbreviated NV (nitroveratryl) and in II, the group is abbreviated NVOC (nitroveratryl oxycarbonyl). Although not shown in FIG. C, the bases (adenine, cytosine, and guanine) contain exocyclic NH_2 groups which must be protected during DNA synthesis. Thymine contains no exocyclic NH_2 and therefore requires no protection. The standard protecting groups for these amines are shown below:



Adenine (A)

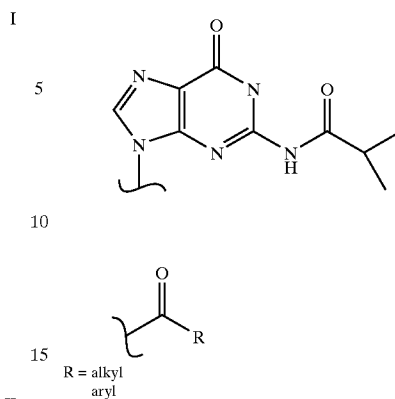


Cytosine (C)

40

-continued

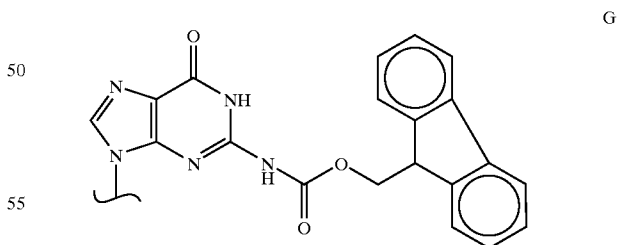
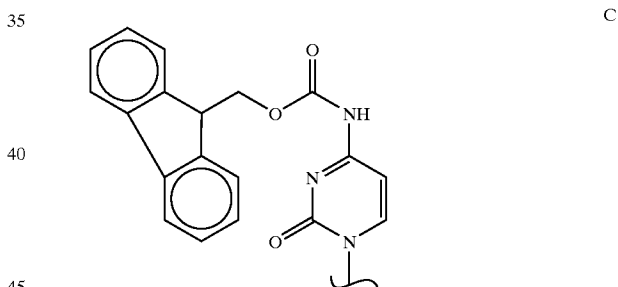
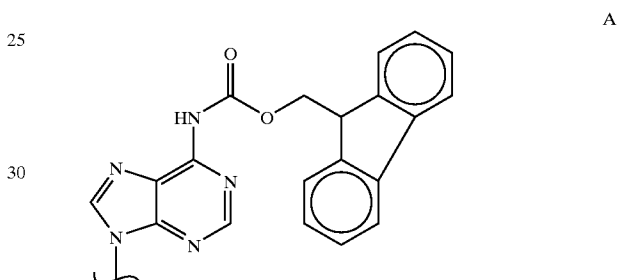
Guanine (G)



II

where R may be alkyl or aryl have been used.

Another type of protecting group Fmoc (9-fluorenyl methoxycarbonyl) is currently being used to protect the exocyclic amines of the three bases:

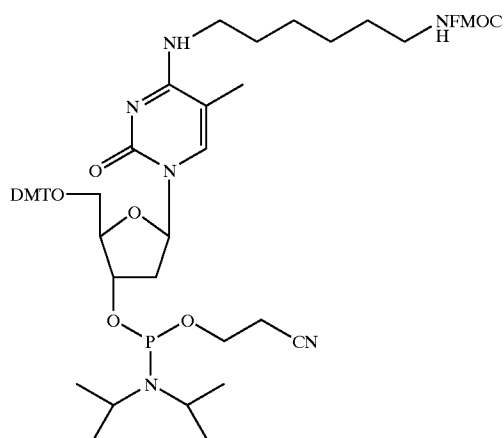


The advantage of the Fmoc group is that it is removed under mild conditions (dilute organic bases) and can be used for all three bases. The amide protecting groups require more harsh conditions to be removed (NH_3/MeOH with heat).

Nucleosides used as 5'-OH probes, useful in verifying correct VLSIPS synthetic function, include, for example, the following:

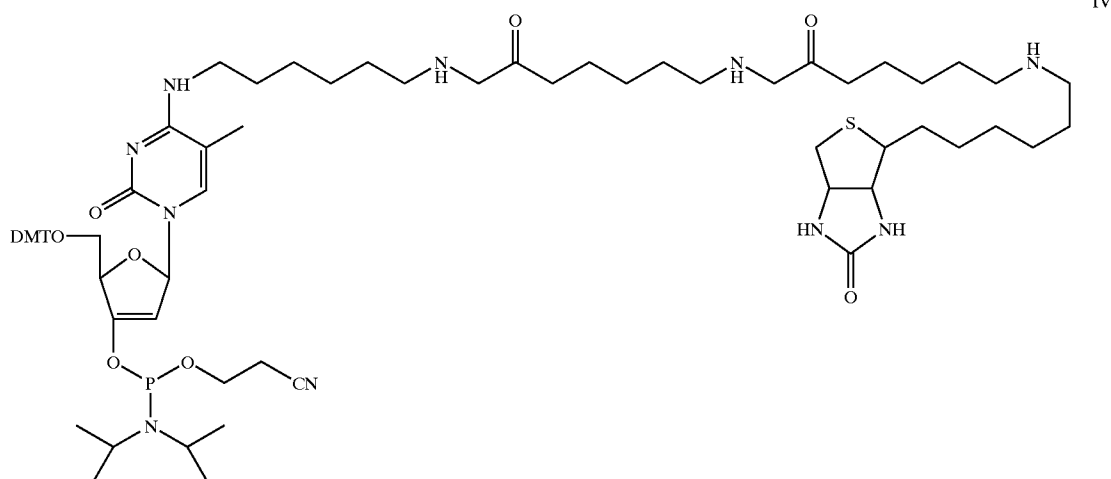
US 6,576,424 B2

41



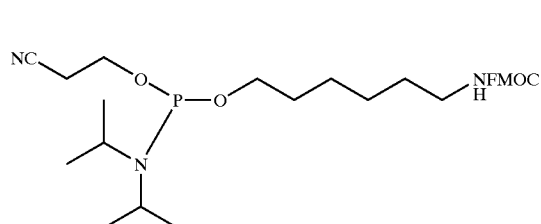
42

III

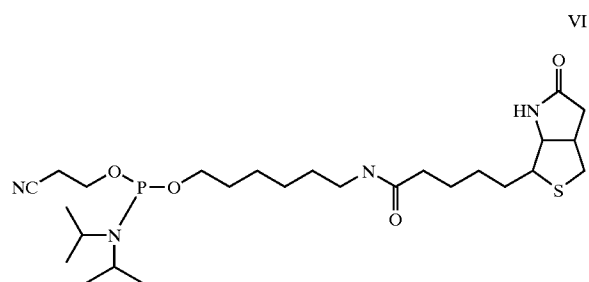


These compounds are used to detect where on a substrate photolysis has occurred by the attachment of either III or V to the newly generated 5'-OH. In the case of III, after the phosphate attachment is made, the substrate is treated with a dilute base to remove the Fmoc group. The resulting amine can be reacted with FITC and the substrate examined by fluorescence microscopy. This indicates the proper generation of a 5'-OH. In the case of compound IV, after the phosphate attachment is made, the substrate is treated with FITC labeled streptavidin and the substrate again may be examined by fluorescence microscopy. Other probes, although not nucleoside based, have included the following:

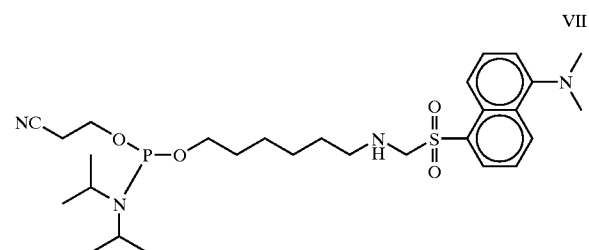
-continued



V



VI

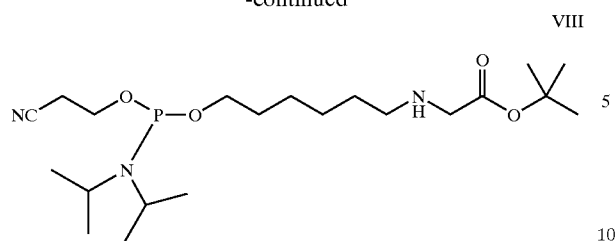


VII

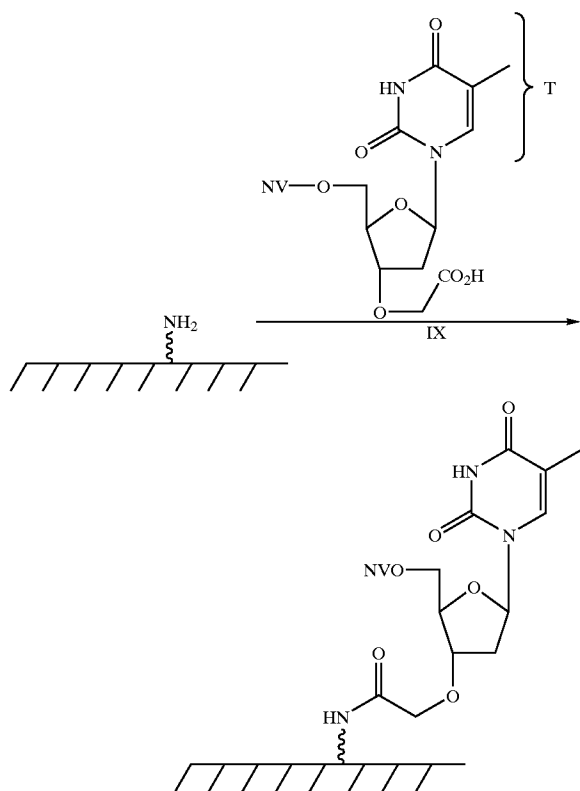
US 6,576,424 B2

43

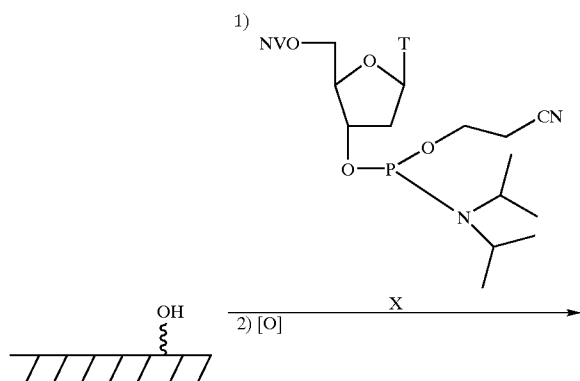
-continued



The method of attachment of the first nucleoside to the surface of the substrate depends on the functionality of the groups at the substrate surface. If the surface is amine functionalized, an amide bond is made (see example below).

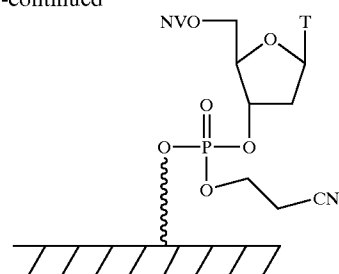


If the surface is hydroxy functionalized, a phosphate bond is made (see example below):



44

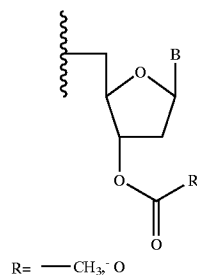
-continued



In both cases, the thymidine example is illustrated, but any one of the four phosphoramidite activated nucleosides can be used in the first step.

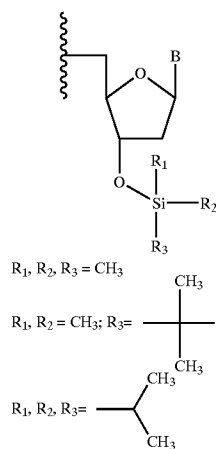
Photolysis of the photolabile group NV or NVOC on the 5' positions of the nucleosides is carried out at ~362 nm with an intensity of 14 mW/cm² for 10 minutes with the substrate side (side containing the photolabile group) immersed in dioxane. After the coupling of the next nucleoside is complete, the photolysis is repeated followed by another coupling until the desired oligomer is obtained.

One of the most common 3'-O-protecting groups is the ester, in particular the acetate:



The groups can be removed by mild base treatment 0.1N NaOH/MeOH or K₂CO₃/H₂O/MeOH.

Another group used most often is the silyl ether:

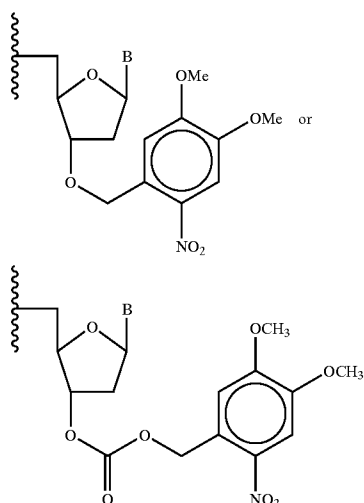


These groups can be removed by neutral conditions using 1 M tetra-n-butylammonium fluoride in THF or under acid conditions.

With respect to photodeprotection, the nitroveratryl group could also be used to protect the 3'-position.

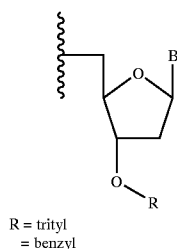
US 6,576,424 B2

45



Here, light (photolysis) would be used to remove these protecting groups.

A variety of others can also be used in the protection of the 3'-O-position:



Removal of these groups usually involves acid or catalytic methods.

Note that corresponding linkages and photoblocked amino acids are described in detail in Ser. No. 07/624,120, now abandoned, which is hereby incorporated herein by reference.

Although the specificity of interactions at particular locations will usually be homogeneous due to a homogeneous polymer being synthesized at each defined location, for certain purposes, it may be useful to have mixed polymers with a commensurate mixed collection of interactions occurring at specific defined locations, or degeneracy reducing analogues, which have been discussed above and show broad specificity in binding. Then, a positive interaction signal may result from any of a number of sequences contained therein.

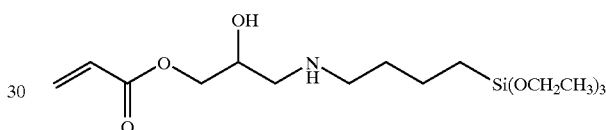
As an alternative method of generating a matrix pattern on a substrate, preformed polymers may be individually attached at particular sites on the substrate. This may be performed by individually attaching reagents one at a time to specific positions on the matrix, a process which may be automated. See, e.g., Ser. No. 07/435,316, now abandoned, and Barrett et al. (1993) U.S. Pat. No. 5,252,743. Another way of generating a positionally defined matrix pattern on a substrate is to have individually specific reagents which interact with each specific position on the substrate. For example, oligonucleotides may be synthesized at defined locations on the substrate. Then the substrate would have on its surface a plurality of regions having homogeneous oligonucleotides attached at each position.

46

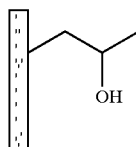
In particular, at least four different substrate preparation procedures are available for treating a substrate surface. They are the standard VLSIPS™ Technology method, polymeric substrates, Durapore™, and synthetic beads or fibers. The treatment labeled "standard VLSIPS™ Technology" method is described in Ser. No. 07/624,120, now abandoned, and involves applying amino-propyltriethoxysilane to a glass surface.

The polymeric substrate approach involves either of two ways of generating a polymeric substrate. The first uses a high concentration of aminopropyltriethoxysilane (2–20%) in an aqueous ethanol solution (95%). This allows the silane compound to polymerize both in solution and on the substrate surface, which provides a high density of amines on the surface of the glass. This density is contrasted with the standard VLSIPS method. This polymeric method allows for the deposition on the substrate surface of a monolayer due to the anhydrous method used with the aforementioned silane.

The second polymeric method involves either the coating or covalent binding of an appropriate acrylic acid polymer onto the substrate surface. In particular, e.g., in DNA synthesis, a monomer such as a hydroxypropylacrylate is used to generate a high density of hydroxyl groups on the substrate surface, allowing for the formation of phosphate bonds. An example of such a compound is shown:



The method using a Durapore™ membrane (Millipore) consists of a polyvinylidene difluoride coating with crosslinked polyhydroxypropyl acrylate [PVDF-HPA]:



Here the building up of, e.g., a DNA oligomer, can be started immediately since phosphate bonds to the surface can be accomplished in the first step with no need for modification. A nucleotide dimer (5'-C-T-3') has been successfully made on this substrate.

The fourth method utilizes synthetic beads or fibers. This would use another substrate, such as a teflon copolymer graft bead or fiber, which is covalently coated with an organic layer (hydrophilic) terminating in hydroxyl sites (commercially available from Molecular Biosystems, Inc.) This would offer the same advantage as the Durapore™ membrane, allowing for immediate phosphate linkages, but would give additional contour by the 3-dimensional growth of oligomers.

A matrix pattern of new reagents may be targeted to each specific oligonucleotide position by attaching a complementary oligonucleotide to which the substrate bound form is complementary. For instance, a number of regions may have homogeneous oligonucleotides synthesized at various locations. Oligonucleotide sequences complementary to each of these can be individually generated and linked to a particular specific reagents. Often these specific reagents will be antibodies. As each of these is specific for finding its

US 6,576,424 B2

47

complementary oligonucleotide, each of the specific reagents will bind through the oligonucleotide to the appropriate matrix position. A single step having a combination of different specific reagents being attached specifically to a particular oligonucleotide will thereby bind to its complement at the defined matrix position. The oligonucleotides will typically then be covalently attached, using, e.g., an acridine dye, for photocrosslinking. Psoralen is a commonly used acridine dye for photocrosslinking purposes, see, e.g., Song et al. (1979) *Photochem. Photobiol.* 29:1177-1197; Cimino et al. (1985) *Ann. Rev. Biochem.* 54:1151-1193; Parsons (1980) *Photochem. Photobiol.* 32:813-821; and Dattagupta et al. (1985) U.S. Pat. No. 4,542,102, and (1987) U.S. Pat. No. 4,713,326; each of which is hereby incorporated herein by reference. This method allows a single attachment manipulation to attach all of the specific reagents to the matrix at defined positions and results in the specific reagents being homogeneously located at defined positions. In many embodiments, the specific reagents will be antibodies.

In an alternative embodiment, antibody molecules may be used to specifically direct binding to defined positions on a substrate. The VLSIPS technology may be used to generate specific epitopes at each position on the substrate. Antibody molecules having specificity of interaction may be used to attach oligonucleotides, thereby avoiding the interference of internal polynucleotide sequences from binding to the substrate complementary oligonucleotides. In fact, the specificity of interaction for positional targeting may be achieved by use of nucleotide analogues which do not interact with the natural nucleotides. For example, other synthetic nucleotides have been made which undergo base pairing, thereby providing the specificity of targeting, but the synthetic nucleotides also do not interact with the natural biological nucleotides. Thus, synthetic oligonucleotides would be useful for attachment to biological nucleotides and specific targeting. Moreover, the VLSIPS synthetic processes would be useful in generating the VLSIPS substrate, and standard oligonucleotide synthesis could be applied, with minor modifications, to produce the complementary sequences which would be attached to other specific reagents.

D. Surface Immobilization

1. Caged Biotin

An alternative method of attaching reagents in a positionally defined matrix pattern is to use a caged biotin system. See Barrett et al. (1993) U.S. Pat. No. 5,252,743, which is hereby incorporated herein by reference, for additional details on the chemistry and application of caged biotin embodiments. In short, the caged biotin has a photosensitive blocking moiety which prevents the combination of avidin to biotin. At positions where the photo-lithographic process has removed the blocking group, high affinity biotin sites are generated. Thus, by a sequential series of photolithographic deblocking steps interspersed with exposure of those regions to appropriate biotin containing reagents, only those locations where the deblocking takes place will form an avidin-biotin interaction. Because the avidin-biotin binding is very tight, this will usually be virtually irreversible binding.

2. Crosslinked Interactions

The surface immobilization may also take place by photocrosslinking of defined oligonucleotides linked to specific reagents. After hybridization of the complementary oligonucleotides, the oligonucleotides may be crosslinked by a reagent by psoralen or another similar type of acridine dye. Other useful cross linking reagents are described in Dattagupta et al. (1985) U.S. Pat. No. 4,542,102, and (1987) U.S. Pat. No. 4,713,326.

48

In another embodiment, colony or phage plaque transfer of biological polymers may be transferred directly onto a silicon substrate. For example, a colony plate may be transferred onto a substrate having a generic oligonucleotide sequence which hybridizes to another generic complementary sequence contained on all of the vectors into which inserts are cloned. This will specifically only bind those molecules which are actually contained in the vectors containing the desired complementary sequence. This immobilization allows for producing a matrix onto which a sequence specific reagent can bind, or for other purposes. In a further embodiment, a plurality of different vectors each having a specific oligonucleotide attached to the vector may be specifically attached to particular regions on a matrix having a complementary oligonucleotide attached thereto.

VIII. HYBRIDIZATION/SPECIFIC INTERACTION

A. General

As discussed previously in the VLSIPS™ Technology parent applications, the VLSIPS™ technology substrates may be used for screening for specific interactions with sequence specific targets or probes.

In addition, the availability of substrates having the entire repertoire of possible sequences of a defined length opens up the possibility of sequencing by hybridization. This sequence may be de novo determination of an unknown sequence, particularly of nucleic acid, verification of a sequence determined by another method, or an investigation of changes in a previously sequenced gene, locating and identifying specific changes. For example, often Maxam and Gilbert sequencing techniques are applied to sequences which have been determined by Sanger and Coulson. Each of those sequencing technologies have problems with resolving particular types of sequences. Sequencing by hybridization may serve as a third and independent method for verifying other sequencing techniques. See, e.g., (1988) *Science* 242:1245.

In addition, the ability to provide a large repertoire of particular sequences allows use of short subsequences and hybridization as a means to fingerprint a sample. This may be used in a nucleic acid, as well as other polymer embodiments. For example, fingerprinting to a high degree of specificity of sequence matching may be used for identifying highly similar samples, e.g., those exhibiting high homology to the selected probes. This may provide a means for determining classifications of particular sequences. This should allow determination of whether particular genomes of bacteria, phage, or even higher cells might be related to one another.

In addition, fingerprinting may be used to identify an individual source of biological sample. See, e.g., Lander, E. (1989) *Nature*, 339:501-505, and references therein. For example, a DNA fingerprint may be used to determine whether a genetic sample arose from another individual. This would be particularly useful in various sorts of forensic tests to determine, e.g., paternity or sources of blood samples. Significant detail on the particulars of genetic fingerprinting for identification purposes are described in, e.g., Morris et al. (1989) "Biostatistical evolution of evidence from continuous allele frequency distribution DNA probes in reference to disputed paternity of identity," *J. Forensic Science* 34:1311-1317; and Neufeld et al. (1990) *Scientific American* 262:46-53; each of which is hereby incorporated herein by reference.

In another embodiment, a fingerprinting-like procedure may be used for classifying cell types by analyzing a pattern of specific nucleic acids present in the cell. A series of

US 6,576,424 B2

49

antibodies may be used to identify cell markers, e.g., proteins, usually on the cell surface, but intracellular markers may also be used. Antigens which are extracellularly expressed are preferred so cell lysis is unnecessary in the screening, but intracellular markers may also be useful. The markers will usually be proteins, but may be nucleic acids, lipids, metabolites, carbohydrates, or other cellular components. See, e.g., Winkelgren, I. (1990) *Science News* 136:234–237, which indicates extracellular DNA may be common, and suggesting that such might be characteristic of cell types, stage, or physiology. This may also be useful in defining the temporal stage of development of cells, e.g., stem cells or other cells which undergo temporal changes in development. For example, the stage of a cell, or group of cells, may be tested or defined by isolating a sample of mRNA from the population and testing to see what sequences are present in messenger populations. Direct samples, or amplified samples, may be used. Where particular mRNA or other nucleic acid sequences may be characteristic of or shown to be characteristic of particular developmental stages, physiological states, or other conditions, this fingerprinting method may define them. Similar sorts of fingerprinting may be used for determining T-cell classes or perhaps even to generate classification schemes for such proteins as major histocompatibility complex antigens. Thus, the ability to make these substrates allows both the generation of reagents which will be used for defining subclasses or classes of cells or other biological materials, but also provides the mechanisms for selecting those cells which may be found in defined population groups.

In addition to cell classification defined by such a combination of properties, typically expression of extracellular antigens, the present invention also provides the means for isolating homogeneous population of cells. Once the antigenic determinants which define a cell class have been identified, these antigens may be used in a sequential selection process to isolate only those cells which exhibit the combination of defining structural properties.

The present invention may also be used for mapping sequences within a larger segment. This may be performed by at least two methods, particularly in reference to nucleic acids. Often, enormous segments of DNA are subcloned into a large plurality of subsequences. Ordering these subsequences may be important in determining the overlaps of sequences upon nucleotide determinations. Mapping may be performed by immobilizing particularly large segments onto a matrix using the VLSIPS™ Technology. Alternatively, sequences may be ordered by virtue of subsequences shared by overlapping segments. See, e.g., Craig et al. (1990) *Nuc. Acids Res.* 18:2653–2660; Michiels et al. (1987) *CABIOS* 3:203–210; and Olson et al. (1986) *Proc. Natl. Acad. Sci. USA* 83:7826–7830.

B. Important Parameters

The extent of specific interaction between reagents immobilized to the VLSIPS™ Technology substrate and another sequence specific reagent may be modified by the conditions of the interaction. Sequencing embodiments typically require high fidelity hybridization and the ability to discriminate perfect matching from imperfect matching. Fingerprinting and mapping embodiments may be performed using less stringent conditions, depending upon the circumstances.

For example, the specificity of antibody/antigen interaction may depend upon such parameters as pH, salt concentration, ionic composition, solvent composition, detergent composition and concentration, and chaotropic agent concentration. See, e.g., Harlow and Lane (1988) *Antibodies: A Laboratory Manual*, Cold Spring Harbor

50

Press, New York. By careful control of these parameters, the affinity of binding may be mapped across different sequences.

In a nucleic acid hybridization embodiment, the specificity and kinetics of hybridization have been described in detail by, e.g., Wetmur and Davidson (1968) *J. Mol. Biol.*, 31:349–370, Britten and Kohne (1968) *Science* 161:529–530, and Kanehisa, (1984) *Nuc. Acids Res.* 12:203–213, each of which is hereby incorporated herein by reference. Parameters which are well known to affect specificity and kinetics of reaction include salt conditions, ionic composition of the solvent, hybridization temperature, length of oligonucleotide matching sequences, guanine and cytosine (GC) content, presence of hybridization accelerators, pH, specific bases found in the matching sequences, solvent conditions, and addition of organic solvents.

In particular, the salt conditions required for driving highly mismatched sequences to completion typically include a high salt concentration. The typical salt used is sodium chloride (NaCl), however, other ionic salts may be utilized, e.g., KCl. Depending on the desired stringency hybridization, the salt concentration will often be less than about 3 molar, more often less than 2.5 molar, usually less than about 2 molar, and more usually less than about 1.5 molar. For applications directed towards higher stringency matching, the salt concentrations would typically be lower. Ordinary high stringency conditions will utilize salt concentration of less than about 1 molar, more often less than about 750 millimolar, usually less than about 500 millimolar, and may be as low as about 250 or 150 millimolar.

The kinetics of hybridization and the stringency of hybridization both depend upon the temperature at which the hybridization is performed and the temperature at which the washing steps are performed. Temperatures at which steps for low stringency hybridization are desired would typically be lower temperatures, e.g., ordinarily at least about 15° C., more ordinarily at least about 20° C., usually at least about 25° C., and more usually at least about 30° C. For those applications requiring high stringency hybridization, or fidelity of hybridization and sequence matching, temperatures at which hybridization and washing steps are performed would typically be high. For example, temperatures in excess of about 35° C. would often be used, more often in excess of about 40° C., usually at least about 45° C., and occasionally even temperatures as high as about 50° C. or 60° C. or more. Of course, the hybridization of oligonucleotides may be disrupted by even higher temperatures. Thus, for stripping of targets from substrates, as discussed below, temperatures as high as 80° C., or even higher may be used.

The base composition of the specific oligonucleotides involved in hybridization affects the temperature of melting, and the stability of hybridization as discussed in the above references. However, the bias of GC rich sequences to hybridize faster and retain stability at higher temperatures can be compensated for by the inclusion in the hybridization incubation or wash steps of various buffers. Sample buffers which accomplish this result include the triethly- and trimethyl ammonium buffers. See, e.g., Wood et al. (1987) *Proc. Natl. Acad. Sci. USA*, 82:1585–1588, and Khrapko, K. et al. (1989) *FEBS Letters* 256:118–122.

The rate of hybridization can also be affected by the inclusion of particular hybridization accelerators. These hybridization accelerators include the volume exclusion agents characterized by dextran sulfate, or polyethylene glycol (PEG). Dextran sulfate is typically included at a concentration of between 1% and 40% by weight. The actual

US 6,576,424 B2

51

concentration selected depends upon the application, but typically a faster hybridization is desired in which the concentration is optimized for the system in question. Dextran sulfate is often included at a concentration of between 0.5% and 2% by weight or dextran sulfate at a concentration between about 0.5% and 5%. Alternatively, proteins which accelerate hybridization may be added, e.g., the recA protein found in *E. coli* or other homologous proteins.

With respect to those embodiments where specific reagents are not oligonucleotides, the conditions of specific interaction would depend on the affinity of binding between the specific reagent and its target. Typically parameters which would be of particular importance would be pH, salt concentration anion and cation compositions, buffer concentration, organic solvent inclusion, detergent concentration, and inclusion of such reagents such as chaotropic agents. In particular, the affinity of binding may be tested over a variety of conditions by multiple washes and repeat scans or by using reagents with differences in binding affinity to determine which reagents bind or do not bind under the selected binding and washing conditions. The spectrum of binding affinities may provide an additional dimension of information which may be very useful in identification purposes and mapping.

Of course, the specific hybridization conditions will be selected to correspond to a discriminatory condition which provides a positive signal where desired but fails to show a positive signal at affinities where interaction is not desired. This may be determined by a number of titration steps or with a number of controls which will be run during the hybridization and/or washing steps to determine at what point the hybridization conditions have reached the stage of desired specificity.

IX. DETECTION METHODS

Methods for detection depend upon the label selected. The criteria for selecting an appropriate label are discussed below, however, a fluorescent label is preferred because of its extreme sensitivity and simplicity. Standard labeling procedures are used to determine the positions where interactions between a sequence and a reagent take place. For example, if a target sequence is labeled and exposed to a matrix of different probes, only those locations where probes do interact with the target will exhibit any signal. Alternatively, other methods may be used to scan the matrix to determine where interaction takes place. Of course, the spectrum of interactions may be determined in a temporal manner by repeated scans of interactions which occur at each of a multiplicity of conditions. However, instead of testing each individual interaction separately, a multiplicity of sequence interactions may be simultaneously determined on a matrix.

A. Labeling Techniques

The target polynucleotide may be labeled by any of a number of convenient detectable markers. A fluorescent label is preferred because it provides a very strong signal with low background. It is also optically detectable at high resolution and sensitivity through a quick scanning procedure. Other potential labeling moieties include, radioisotopes, chemiluminescent compounds, labeled binding proteins, heavy metal atoms, spectroscopic markers, magnetic labels, and linked enzymes.

Another method for labeling may bypass any label of the target sequence. The target may be exposed to the probes, and a double strand hybrid is formed at those positions only. Addition of a double strand specific reagent will detect where hybridization takes place. An intercalative dye such as

52

ethidium bromide may be used as long as the probes themselves do not fold back on themselves to a significant extent forming hairpin loops. See, e.g., Sheldon et al. (1986) U.S. Pat. No. 4,582,789. However, the length of the hairpin loops in short oligonucleotide probes would typically be insufficient to form a stable duplex.

In another embodiment, different targets may be simultaneously sequenced where each target has a different label. For instance, one target could have a green fluorescent label and a second target could have a red fluorescent label. The scanning step will distinguish sites of binding of the red label from those binding the green fluorescent label. Each sequence can be analyzed independently from one another.

Suitable chromogens will include molecules and compounds which absorb light in a distinctive range of wavelengths so that a color may be observed, or emit light when irradiated with radiation of a particular wave length or wave length range, e.g., fluorescers. Biliproteins, e.g., phycoerythrin, may also serve as labels.

A wide variety of suitable dyes are available, being primarily chosen to provide an intense color with minimal absorption by their surroundings. Illustrative dye types include quinoline dyes, triarylmethane dyes, acridine dyes, alizarine dyes, phthaleins, insect dyes, azo dyes, anthraquinoid dyes, cyanine dyes, phenazathionium dyes, and phenazonium dyes.

A wide variety of fluorescers may be employed either by themselves or in conjunction with quencher molecules. Fluorescers of interest fall into a variety of categories having certain primary functionalities. These primary functionalities include 1- and 2-aminonaphthalene, p,p'-diaminostilbenes, pyrenes, quaternary phenanthridine salts, 9-aminoacridines, p,p'-diaminobenzophenone imines, anthracenes, oxacarbocyanine, merocyanine, 3-aminoequilenin, perylene, bis-benzoxazole, bis-p-oxazolyl benzene, 1,2-benzophenazin, retinol, bis-3-aminopyridinium salts, hellebrigenin, tetracycline, sterophenol, benzimidazoalylphenylamine, 2-oxo-3-chromen, indole, xanthen, 7-hydroxycoumarin, phenoxazine, salicylate, strophanthidin, porphyrins, triarylmethanes and flavin. Individual fluorescent compounds which have functionalities for linking or which can be modified to incorporate such functionalities include, e.g., dansyl chloride; fluoresceins such as 3,6-dihydroxy-9-phenylxanthidrol; rhodamineisothiocyanate; N-phenyl 1-amino-8-sulfonatonaphthalene; N-phenyl 2-amino-6-sulfonatonaphthalene; 4-acetamido-4-isothiocyanato-stilbene-2,2'-disulfonic acid; pyrene-3-sulfonic acid; 2-toluidinonaphthalene-6-sulfonate; N-phenyl, N-methyl 2-aminoaphthalene-6-sulfonate; ethidium bromide; stebrine; aurochrome-0,2-(9'-anthroyl)palmitate; dansyl phosphatidylethanolamine; N,N'-dioctadecyl oxacarbocyanine; N,N'-dihexyl oxacarbocyanine; merocyanine, 4-(3'pyrenyl) butyrate; d-3-aminodesoxy-equilenin; 12-(9'-anthroyl) stearate; 2-methylanthracene; 9-vinylanthracene; 2,2'-(vinylene-p-phenylene)bisbenzoxazole; p-bis[2-(4-methyl-5-phenyl-oxazolyl)]benzene; 6-dimethylamino-1,2-benzophenazin; retinol; bis (3'-aminopyridinium) 1,10-decandiyl diiodide; sulfonaphthylhydrazine of hellebrigenin; chlorotetracycline; N-(7-dimethylamino-4-methyl-2-oxo-3-chromenyl)maleimide; N-[p-(2-benzimidazolyl)-phenyl] maleimide; N-(4-fluoranthyl)maleimide; bis(homovanillic acid); resazurin; 4-chloro-7-nitro-2,1,3-benzooxadiazole; merocyanine 540; resorufin; rose bengal; and 2,4-diphenyl-3(2H)-furanone.

Desirably, fluorescers should absorb light above about 300 nm, preferably about 350 nm, and more preferably

US 6,576,424 B2

53

above about 400 nm, usually emitting at wavelengths greater than about 10 nm higher than the wavelength of the light absorbed. It should be noted that the absorption and emission characteristics of the bound dye may differ from the unbound dye. Therefore, when referring to the various wavelength ranges and characteristics of the dyes, it is intended to indicate the dyes as employed and not the dye which is unconjugated and characterized in an arbitrary solvent.

Fluorescers are generally preferred because by irradiating a fluorescer with light, one can obtain a plurality of emissions. Thus, a single label can provide for a plurality of measurable events.

Detectable signal may also be provided by chemiluminescent and bioluminescent sources. Chemiluminescent sources include a compound which becomes electronically excited by a chemical reaction and may then emit light which serves as the detectible signal or donates energy to a fluorescent acceptor. A diverse number of families of compounds have been found to provide chemiluminescence under a variety of conditions. One family of compounds is 2,3-dihydro-1,4-phthalazinedione. The most popular compound is luminol, which is the 5-amino compound. Other members of the family include the 5-amino-6,7,8-trimethoxy- and the dimethylamino[ca]benz analog. These compounds can be made to luminesce with alkaline hydrogen peroxide or calcium hypochlorite and base. Another family of compounds is the 2,4,5-triphenylimidazoles, with lophine as the common name for the parent product. Chemiluminescent analogs include para-dimethylamino and -methoxy substituents. Chemiluminescence may also be obtained with oxalates, usually oxalyl active esters, e.g., p-nitrophenyl and a peroxide, e.g., hydrogen peroxide, under basic conditions. Alternatively, luciferins may be used in conjunction with luciferase or lucigenins to provide bioluminescence.

Spin labels are provided by reporter molecules with an unpaired electron spin which can be detected by electron spin resonance (ESR) spectroscopy. Exemplary spin labels include organic free radicals, transitional metal complexes, particularly vanadium, copper, iron, and manganese, and the like. Exemplary spin labels include nitroxide free radicals.

B. Scanning System

With the automated detection apparatus, the correlation of specific positional labeling is converted to the presence on the target of sequences for which the reagents have specificity of interaction. Thus, the positional information is directly converted to a database indicating what sequence interactions have occurred. For example, in a nucleic acid hybridization application, the sequences which have interacted between the substrate matrix and the target molecule can be directly listed from the positional information. The detection system used is described in Pirrung et al. (1992) U.S. Pat. No. 5,143,854; and Ser. No. 07/624,120, now abandoned. Although the detection described therein is a fluorescence detector, the detector may be replaced by a spectroscopic or other detector. The scanning system may make use of a moving detector relative to a fixed substrate, a fixed detector with a moving substrate, or a combination. Alternatively, mirrors or other apparatus can be used to transfer the signal directly to the detector. See, e.g., Ser. No. 07/624,120, now abandoned, which is hereby incorporated herein by reference.

The detection method will typically also incorporate some signal processing to determine whether the signal at a particular matrix position is a true positive or may be a spurious signal. For example, a signal from a region which

54

has actual positive signal may tend to spread over and provide a positive signal in an adjacent region which actually should not have one. This may occur, e.g., where the scanning system is not properly discriminating with sufficiently high resolution in its pixel density to separate the two regions. Thus, the signal over the spatial region may be evaluated pixel by pixel to determine the locations and the actual extent of positive signal. A true positive signal should, in theory, show a uniform signal at each pixel location. Thus, processing by plotting number of pixels with actual signal intensity should have a clearly uniform signal intensity. Regions where the signal intensities show a fairly wide dispersion, may be particularly suspect and the scanning system may be programmed to more carefully scan those positions.

In another embodiment, as the sequence of a target is determined at a particular location, the overlap for the sequence would necessarily have a known sequence. Thus, the system can compare the possibilities for the next adjacent position and look at these in comparison with each other. Typically, only one of the possible adjacent sequences should give a positive signal and the system might be programmed to compare each of these possibilities and select that one which gives a strong positive. In this way, the system can also simultaneously provide some means of measuring the reliability of the determination by indicating what the average signal to background ratio actually is.

More sophisticated signal processing techniques can be applied to the initial determination of whether a positive signal exists or not. See, e.g., Ser. No. 07/624,120, now abandoned.

From a listing of those sequences which interact, data analysis may be performed on a series of sequences. For example, in a nucleic acid sequence application, each of the sequences may be analyzed for their overlap regions and the original target sequence may be reconstructed from the collection of specific subsequences obtained therein. Other sorts of analyses for different applications may also be performed, and because the scanning system directly interfaces with a computer the information need not be transferred manually. This provides for the ability to handle large amounts of data with very little human intervention. This, of course, provides significant advantages over manual manipulations. Increased throughput and reproducibility is thereby provided by the automation of a vast majority of steps in any of these applications.

XI. DATA ANALYSIS

A. General

Data analysis will typically involve aligning the proper sequences with their overlaps to determine the target sequence. Although the target "sequence" may not specifically correspond to any specific molecule, especially where the target sequence is broken and fragmented in the sequencing process, the sequence corresponds to a contiguous sequence of the subfragments.

The data analysis can be performed by a computer using an appropriate program. See, e.g., Drmanac, R. et al. (1989) *Genomics* 4:114-128; and a commercially available analysis program available from the Genetic Engineering Center, P.O. Box 794, 11000 Belgrade, Yugoslavia. Although the specific manipulations necessary to reassemble the target sequence from fragments may take many forms, one embodiment uses a sorting program to sort all of the subsequences using a defined hierarchy. The hierarchy need not necessarily correspond to any physical hierarchy, but provides a means to determine, in order, which subfragments have actually been found in the target sequence. In this

US 6,576,424 B2

55

manner, overlaps can be checked and found directly rather than having to search throughout the entire set after each selection process. For example, where the oligonucleotide probes are 10-mers, the first 9 positions can be sorted. A particular subsequence can be selected as in the examples, to determine where the process starts. As analogous to the theoretical example provided above, the sorting procedure provides the ability to immediately find the position of the subsequence which contains the first 9 positions and can compare whether there exists more than 1 subsequence during the first 9 positions. In fact, the computer can easily generate all of the possible target sequences which contain given combination of subsequences. Typically there will be only one, but in various situations, there will be more.

An exemplary flow chart for a sequencing program is provided in FIG. 1. In general terms, the program provides for automated scanning of the substrate to determine the positions of probe and target interaction. Simple processing of the intensity of the signal may be incorporated to filter out clearly spurious signals. The positions with positive interaction are correlated with the sequence specificity of specific matrix positions, to generate the set of matching subsequences. This information is further correlated with other target sequence information, e.g., restriction fragment analysis. The sequences are then aligned using overlap data, thereby leading to possible corresponding target sequences which will, optimally, correspond to a single target sequence.

B. Hardware

A variety of computer systems may be used to run a sequencing program. The program may be written to provide both the detecting and scanning steps together and will typically be dedicated to a particular scanning apparatus. However, the components and functional steps may be separated and the scanning system may provide an output, e.g., through tape or an electronic connection into a separate computer which separately runs the sequencing analysis program. The computer may be any of a number of machines provided by standard computer manufacturers, e.g., IBM compatible machines, Apple™ machines, VAX machines, and others, which may often use a UNIX™ operating system. Of course, the hardware used to run the analysis program will typically determine what programming language would be used.

C. Software

Software would be easily developed by a person of ordinary skill in the programming art, following the flow chart provided, or based upon the input provided and the desired result.

Of course, an exemplary embodiment is a polynucleotide sequence system. However, the theoretical and mathematical manipulations necessary for data analysis of other linear molecules, such as polypeptides, carbohydrates, and various other polymers are conceptually similar. Simple branching polymers will usually also be sequencable using similar technology. However, where there is branching, it may be desired that additional recognition reagents be used to determine the nature and location of branches. This can easily be provided by use of appropriate specific reagents which would be generated by methods similar to those used to produce specific reagents for linear polymers.

XII. SUBSTRATE REUSE

Where a substrate is made with specific reagents that are relatively insensitive to the handling and processing steps involved in a single cycle of use, the substrate may often be reused. The target molecules are usually stripped off of the

56

solid phase specific recognition molecules. Of course, it is preferred that the manipulations and conditions be selected as to be mild and to not affect the substrate. For example, if a substrate is acid labile, a neutral pH would be preferred in all handling steps. Similar sensitivities would be carefully respected where recycling is desired.

A. Removal of Label

Typically for a recycling, the previously attached specific interaction would be disrupted and removed. This will typically involve exposing the substrate to conditions under which the interaction between probe and target is disrupted. Alternatively, it may be exposed to conditions where the target is destroyed. For example, where the probes are oligonucleotides and the target is a polynucleotide, a heating and low salt wash will often be sufficient to disrupt the interactions. Additional reagents may be added such as detergents, and organic or inorganic solvents which disrupt the interaction between the specific reagents and target. In an embodiment where the specific reagents are antibodies, the substrate may be exposed to a gentle detergent which will denature the specific binding between the antibody and its target. The conditions are selected to avoid severe disruption or destruction of the structure of the antibody and to maintain the specificity of the antibody binding site. Conditions with specific pH, detergent concentration, salt concentration, ionic concentration, and other parameters may be selected which disrupt the specific interactions.

B. Storage and Preservation

As indicated above, the matrix will typically be maintained under conditions where the matrix itself and the linkages and specific reagents are preserved. Various specific preservatives may be added which prevent degradation. For example, if the reagents are acid or base labile, a neutral pH buffer will typically be added. It is also desired to avoid destruction of the matrix by growth of organisms which may destroy organic reagents attached thereto. For this reason, a preservative such as cyanide or azide may be added. However, the chemical preservative should also be selected to preserve the chemical nature of the linkages and other components of the substrate. Typically, a detergent may also be included.

C. Processes to Avoid Degradation of Oligomers

In particular, a substrate comprising a large number of oligomers will be treated in a fashion which is known to maintain the quality and integrity of oligonucleotides. These include storing the substrate in a carefully controlled environment under conditions of lower temperature, cation depletion (EDTA and EGTA), sterile conditions, and inert argon or nitrogen atmosphere.

XIII. INTEGRATED SEQUENCING STRATEGY

A. Initial Mapping Strategy

As indicated above, although the VLSIPS™ technology may be applied to sequencing embodiments, it is often useful to integrate other concepts to simplify the sequencing. For example, nucleic acids may be easily sequenced by careful selection of the vectors and hosts used for amplifying and generating the specific target sequences. For example, it may be desired to use specific vectors which have been designed to interact most efficiently with the VLSIPS substrate. This is also important in fingerprinting and mapping strategies. For example, vectors may be carefully selected having particular complementary sequences which are designed to attach to a genetic or specific oligomer on the substrate. This is also applicable to situations where it is desired to target particular sequences to specific locations on the matrix.

In one embodiment, unnatural oligomers may be used to target natural probes to specific locations on the VLSIPS

US 6,576,424 B2

57

substrate. In addition, particular probes may be generated for the mapping embodiment which are designed to have specific combinations of characteristics. For example, the construction of a mapping substrate may depend upon use of another automated apparatus which takes clones isolated from a chromosome walk and attaches them individually or in bulk to the VLSIPS substrate.

In another embodiment, a variety of specific vectors having known and particular "targeting" sequences adjacent to the cloning sites may be individually used to clone a selected probe, and the isolated probe will then be targetable to a site on the VLSIPS substrate with a sequence complementary to the "target" sequence.

B. Selection of Smaller Clones

In the fingerprinting and mapping embodiments, the selection of probes may be very important. Significant mathematical analysis may be applied to determine which specific sequences should be used as those probes. Of course, for fingerprinting use, these sequences would be most desired that show significant heterogeneity across the human population. Selection of the specific sequences which would most favorably be utilized will tend to be single copy sequences within the genome.

Various hybridization selection procedures may be applied to select sequences which tend not to be repeated within a genome, and thus would tend to be conserved across individuals. For example, hybridization selections may be made for non-repetitive and single copy sequences. See, e.g., Britten and Kohne (1968) "Repeated Sequences in DNA," *Science* 161:529-540. On the other hand, it may be desired under certain circumstances to use repeated sequences. For example, where a fingerprint may be used to identify or distinguish different species, or where repetitive sequences may be diagnostic of specific species, repetitive sequences may be desired for inclusion in the fingerprinting probes. In either case, the sequencing capability will greatly assist in the selection of appropriate sequences to be used as probes.

Also as indicated above, various means for constructing an appropriate substrate may involve either mechanical or automated procedures. The standard VLSIPS automated procedure involves synthesizing oligonucleotides or short polymers directly on the substrate. In various other embodiments, it is possible to attach separately synthesized reagents onto the matrix in an ordered array. Other circumstances may lend themselves to transfer a pattern from a petri plate onto a solid substrate. Also, there are methods for site specifically directing collections of reagents to specific locations using unnatural nucleotides or equivalent sorts of targeting molecules.

While a brute force manual transfer process may be utilized sequentially for attaching various samples to successive positions, instrumentation for automating such procedures may also be devised. The automated system for performing such would preferably be relatively easily designed and conceptually easily understood.

XIV. COMMERCIAL APPLICATIONS

A. Sequencing

As indicated above, sequencing may be performed either de novo or as a verification of another sequencing method. The present hybridization technology provides the ability to sequence nucleic acids and polynucleotides de novo, or as a means to verify either the Maxam and Gilbert chemical sequencing technique or Sanger and Coulson dideoxy-sequencing techniques. The hybridization method is useful to verify sequencing determined by any other sequencing technique and to closely compare two similar sequences, e.g., to identify and locate sequence differences.

58

Besides polynucleotide sequencing, the present invention also provides means for sequencing other polymers. This includes polypeptides, carbohydrates, synthetic organic polymers, and other polymers. Again, the sequencing may be either verification or de novo.

Of course, sequencing can be very important in many different sorts of environments. For example, it will be useful in determining the genetic sequence of particular markers in various individuals. In addition, polymers may be used as markers or for information containing molecules to encode information. For example, a short polynucleotide sequence may be included in large bulk production samples indicating the manufacturer, date, and location of manufacture of a product. For example, various drugs may be encoded with this information with a small number of molecules in a batch. For example, a pill may have somewhere from 10 to 100 to 1,000 or more very short and small molecules encoding this information. When necessary, this information may be decoded from a sample of the material using a polymerase chain reaction (PCR) or other amplification method. This encoding system may be used to provide the origin of large bulky samples without significantly affecting the properties of those samples. For example, chemical samples may also be encoded by this method thereby providing means for identifying the source and manufacturing details of lots. The origin of bulk hydrocarbon samples may be encoded. Production lots of organic compounds such as benzene or plastics may be encoded with a short molecule polymer. Food stuffs may also be encoded using similar marking molecules. Even toxic waste samples can be encoded determining the source or origin. In this way, proper disposal can be traced or more easily enforced.

Similar sorts of encoding may be provided by fingerprinting-type analysis. Whether the resolution is absolute or less so, the concept of coding information on molecules such as nucleic acids, which can be amplified and later decoded, may be a very useful and important application.

This technology also provides the ability to include markers for origins of biological materials. For example, a patented animal line may be transformed with a particular unnatural sequence which can be traced back to its origin. With a selection of multiple markers, the likelihood could be negligible that a combination of markers would have independently arisen from a source other than the patented or specifically protected source. This technique may provide a means for tracing the actual origin of particular biological materials. Bacteria, plants, and animals will be subject to marking by such encoding sequences.

B. Fingerprinting

As indicated above, fingerprinting technology may also be used for data encryption. Moreover, fingerprinting allows for significant identification of particular individuals. Where the fingerprinting technology is standardized, and used for identification of large numbers of people, related equipment and peripheral processing will be developed to accompany the underlying technology. For example, specific equipment may be developed for automatically taking a biological sample and generating or amplifying the information molecules within the sample to be used in fingerprinting analysis. Moreover, the fingerprinting substrate may be mass produced using particular types of automatic equipment. Synthetic equipment may produce the entire matrix simultaneously by stepwise synthetic methods as provided by the VLSIPS™ technology. The attachment of specific probes onto a substrate may also be automated, e.g., making use of the caged biotin technology. See, e.g., Barrett et al. (1993)

US 6,576,424 B2

59

U.S. Pat. No. 5,252,743. As indicated above, there are automated methods for actually generating the matrix and substrate with distinct sequence reagents positionally located at each of the matrix positions. Where such reagents are, e.g., unnatural amino acids, a targeting function may be utilized which does not interfere with a natural nucleotide functionality.

In addition, peripheral processing may be important and may be dedicated to this specific application. Thus, automated equipment for producing the substrates may be designed, or particular systems which take in a biological sample and output either a computer readout or an encoded instrument, e.g., a card or document which indicates the information and can provide that information to others. An identification having a short magnetic strip with a few million bits may be used to provide individual identification and important medical information useful in a medical emergency.

In fact, data banks may be set up to correlate all of this information of fingerprinting with medical information. This may allow for the determination of correlations between various medical problems and specific DNA sequences. By collating large populations of medical records with genetic information, genetic propensities and genetic susceptibilities to particular medical conditions may be developed. Moreover, with standardization of substrates, the micro encoding data may be also standardized to reproduce the information from a centralized data bank or on an encoding device carried on an individual person. On the other hand, if the fingerprinting procedure is sufficiently quick and routine, every hospital may routinely perform a fingerprinting operation and from that determine many important medical parameters for an individual.

In particular industries, the VLSIPS sequencing, fingerprinting, or mapping technology will be particularly appropriate. As mentioned above, agricultural livestock suppliers may be able to encode and determine whether their particular strains are being used by others. By incorporating particular markers into their genetic stocks, the markers will indicate origin of genetic material. This is applicable to seed producers, livestock producers, and other suppliers of medical or agricultural biological materials.

This may also be useful in identifying individual animals or plants. For example, these markers may be useful in determining whether certain fish return to their original breeding grounds, whether sea turtles always return to their original birthplaces, or to determine the migration patterns and viability of populations of particular endangered species. It would also provide means for tracking the sources of particular animal products. For example, it might be useful for determining the origins of controlled animal substances such as elephant ivory or particular bird populations whose importation or exportation is controlled.

As indicated above, polymers may be used to encode important information on source and batch and supplier. This is described in greater detail, e.g., "Applications of PCR to industrial problems," (1990) in Chemical and Engineering News 68:145, which is hereby incorporated herein by reference. In fact, the synthetic method can be applied to the storage of enormous amounts of information. Small substrates may encode enormous amounts of information, and its recovery will make use of the inherent replication capacity. For example, on regions of $10\mu\text{m}\times 10\mu\text{m}$, 1 cm has 10^6 regions. In theory, the entire human genome could be attached in 1000 nucleotide segments on a 3 cm^2 surface. Genomes of endangered species may be stored on these substrates.

60

Fingerprinting may also be used for genetic tracing or for identifying individuals for forensic science purposes. See, e.g., Morris, J. et al. (1989) "Biostatistical Evaluation of Evidence From Continuous Allele Frequency Distribution DNA Probes in Reference to Disputed Paternity and Identity," *J. Forensic Science* 34:1311-1317, and references provided therein; each of which is hereby incorporated herein by reference.

In addition, the high resolution fingerprinting allows the distinguishability to high resolution of particular samples. As indicated above, new cell classifications may be defined based on combinations of a large number of properties. Similar applications will be found in distinguishing different species of animals or plants. In fact, microbial identification may become dependent on characterization of the genetic content. Tumors or other cells exhibiting abnormal physiology will be detectable by use of the present invention. Also, knowing the genetic fingerprint of a microorganism may provide very useful information on how to treat an infection by such organism.

Modifications of the fingerprint embodiments may be used to diagnose the condition of the organism. For example, a blood sample is presently used for diagnosing any of a number of different physiological conditions. A multi-dimensional fingerprinting method made available by the present invention could become a routine means for diagnosing an enormous number of physiological features simultaneously. This may revolutionize the practice of medicine in providing information on an enormous number of parameters together at one time. In another way, the genetic predisposition may also revolutionize the practice of medicine providing a physician with the ability to predict the likelihood of particular medical conditions arising at any particular moment. It also provides the ability to apply preventive medicine.

The present invention might also find application in use for screening new drugs and new reagents which may be very important in medical diagnosis or other applications. For example, a description of generating a population of monoclonal antibodies with defined specificities may be very useful for producing various drugs or diagnostic reagents.

Also available are kits with the reagents useful for performing sequencing, fingerprinting, and mapping procedures. The kits will have various compartments with the desired necessary reagents, e.g., substrate, labeling reagents for target samples, buffers, and other useful accompanying products.

C. Mapping

The present invention also provides the means for mapping sequences within enormous stretches of sequence. For example, nucleotide sequences may be mapped within enormous chromosome size sequence maps. For example, it would be possible to map a chromosomal location within the chromosome which contains hundreds of millions of nucleotide base pairs. In addition, the mapping and fingerprinting embodiments allow for testing of chromosomal translocations, one of the standard problems for which amniocentesis is performed.

Thus, the present invention provides a powerful tool and the means for performing sequencing, fingerprinting, and mapping functions on polymers. Although most easily and directly applicable to polynucleotides, polypeptides, carbohydrates, and other sorts of molecules can be advantageously utilized using the present technology.

The present invention will be better understood by reference to the following illustrative examples. The following examples are offered by way of illustration and not by way of limitation.

US 6,576,424 B2

61

EXPERIMENTAL

- I. Sequencing
 - A. polynucleotide
 - B. polypeptide
 - C. short peptide
 - 1. Herz antibody identification
- II. Fingerprinting
 - A. polynucleotide fingerprint
 - B. peptide fingerprint
 - C. cell classification scheme
 - D. temporal development scheme
 - 1. developmental antigens
 - 2. developmental mRNA expression
 - E. diagnostic test
 - 1. viral identification
 - 2. bacterial identification
 - 3. other microbiological identifications
 - 4. allergy test (immobilized antigens)
 - F. individual (animal/plant) identification
 - 1. genetic
 - 2. immunological
 - G. genetic screen
 - 1. test alleles with markers
 - 2. amniocentesis
- III. Mapping
 - A. positionally located clones (caged biotin)
 - 1. short probes, long targets
 - 2. long targets, short probes
 - B. positionally defined clones
- IV. Conclusion

Relevant applications whose techniques are incorporated herein by reference are Pirrung, et al., Ser. No. 07/362,901, filed Jun. 7, 1989, now abandoned; Pirrung et al. (1992) U.S. Pat. No. 5,143,854; Barrett, et al., Ser. No. 07/435,316 filed Nov. 13, 1989, now abandoned; Barrett, et al. (1993) U.S. Pat. No. 5,252,743; and commonly assigned and simultaneously filed applications Ser. No. 07/624,120, now abandoned, and Ser. No. 07/626,730.

Also, additional relevant techniques are described, e.g., in Sambrook, J., et al. (1989) *Molecular Cloning: a Laboratory Manual*, 2d Ed., vols 1-3, Cold Spring Harbor Press, N.Y.; Greenstein and Winitz (1961) *Chemistry of the Amino Acids*, Wiley and Sons, New York; Bodzansky, M. (1988) *Peptide Chemistry: a Practical Textbook*, Springer-Verlag, New

62

York; Harlow and Lane (1988) *Antibodies: A Laboratory Manual*, Cold Spring Harbor Press, New York; Glover, D. (ed.) (1987) *DNA Cloning: A Practical Approach*, vols 1-3, IRL Press, Oxford; Bishop and Rawlings (1987) *Nucleic Acid and Protein Sequence Analysis: A Practical Approach*, IRL Press, Oxford; Hames and Higgins (1985) *Nucleic Acid Hybridisation: A Practical Approach*, IRL Press, Oxford; Wu et al. (1989) *Recombinant DNA Methodology*, Academic Press, San Diego; Goding (1986) *Monoclonal Antibodies: Principles and Practice*, (2d ed.), Academic Press, San Diego; Finegold and Barron (1986) *Bailey and Scott's Diagnostic Microbiology*, (7th ed.), Mosby Co., St. Louis; Collins et al. (1989) *Microbiological Methods*, (6th ed.), Butterworth, London; Chaplin and Kennedy (1986) *Carbohydrate Analysis: A Practical Approach*, IRL Press, Oxford; Van Dyke (ed.) (1985) *Bioluminescence and Chemiluminescence: Instruments and Applications*, vol 1, CRC Press, Boca Rotan; and Ausubel et al. (ed.) (1990) *Current Protocols in Molecular Biology*, Greene Publishing and Wiley-Interscience, New York; each of which is hereby incorporated herein by reference.

The following examples are provided to illustrate the efficacy of the inventions herein. All operations were conducted at about ambient temperatures and pressures unless indicated to the contrary.

I. SEQUENCING

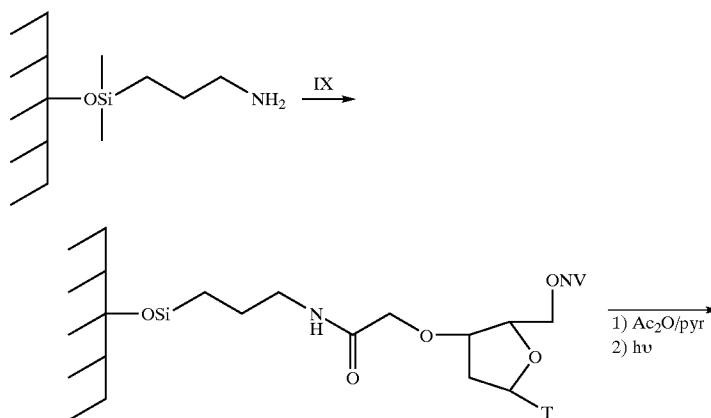
A. Polynucleotide

1. HPLC of the Photolysis of 5'-O-Nitroveratryl-thymidine.

In order to determine the time for photolysis of 5'-O-nitroveratryl thymidine to thymidine a 100 μ M solution of NV-Thym-OH (5'-O-nitroveratryl thymidine) in dioxane was made and ~200 μ l aliquots were irradiated (in a quartz cuvette 1 cm \times 2 mm) at 362.3 nm for 20 sec, 40 sec, 60 sec, 2 min, 5 min, 10 min, 15 min, and 20 min. The resulting irradiated mixtures were then analyzed by HPLC using a Varian MicroPak SP column (C₁₈ analytical) at a flow rate of 1 ml/min and a solvent system of 40% CH₃CN and 60% water. Thymidine has a retention time of 1.2 min and NVO-Thym-OH has a retention time of 2.1 min. It was seen that after 10 min of exposure the deprotection was complete.

2. Preparation and Detection of Thymidine-Cytidine Dimer (FITC)

The reaction is illustrated:

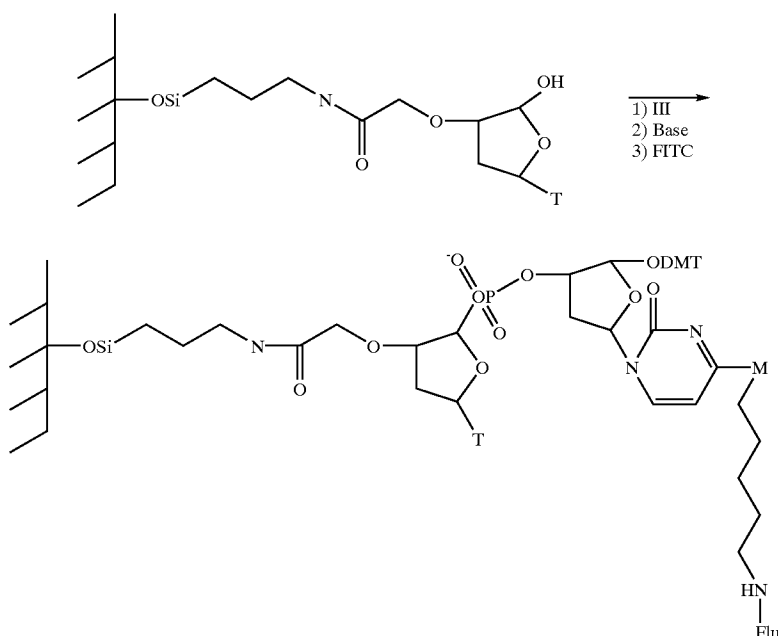


US 6,576,424 B2

63

-continued

64



To an aminopropylated glass slide (standard VLSIPS™ Technology) was added a mixture of the following:

12.2 mg of NVO-Thym-CO₂H (IX)

3.4 mg of HOBT (N-hydroxybenztriazol)

8.8 μ l DIEA (Diisopropylethylamine)

11.1 mg BOP reagent

2.5 ml DMF

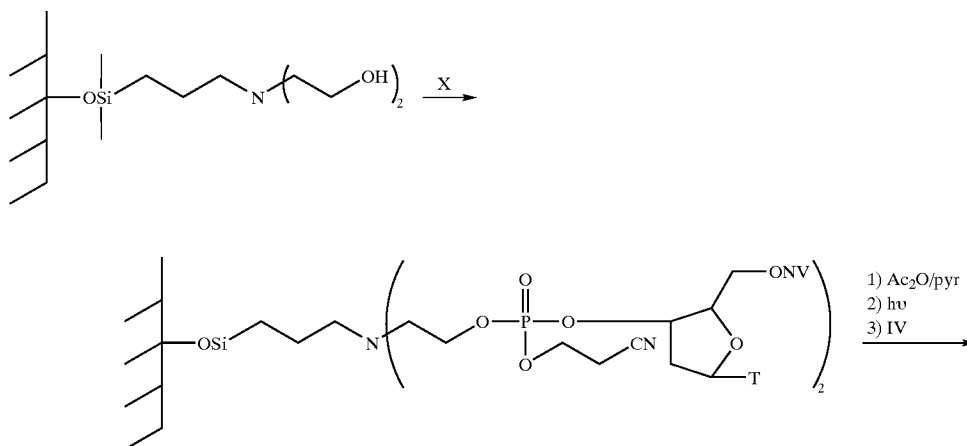
After 2 h coupling time (standard VLSIPS) the plate was washed, acetylated with acetic anhydride/pyridine, washed, dried, and photolyzed in dioxane at 362 nm at 14 mW/cm² for 10 min using a 500 μ m checkerboard mask. The slide was then taken and treated with a mixture of the following:

107 mg of Fmoc-amine modified C (III)

21 mg of tetrazole

1 ml anhydrous CH₃CN

After being treated for approximately 8 min, the slide was washed off with CH₃CN, dried, and oxidized with I₂/H₂O/THF/lutidine for 1 min. The slide was again washed, dried, and treated for 30 min with a 20% solution of DBU in DMF. After thorough rinsing of the slide, it was next exposed to a FITC solution (1 mM fluorescein isothiocyanate [FITC] in DMF) for 50 min, then washed, dried, and examined by fluorescence microscopy. This reaction is illustrated:

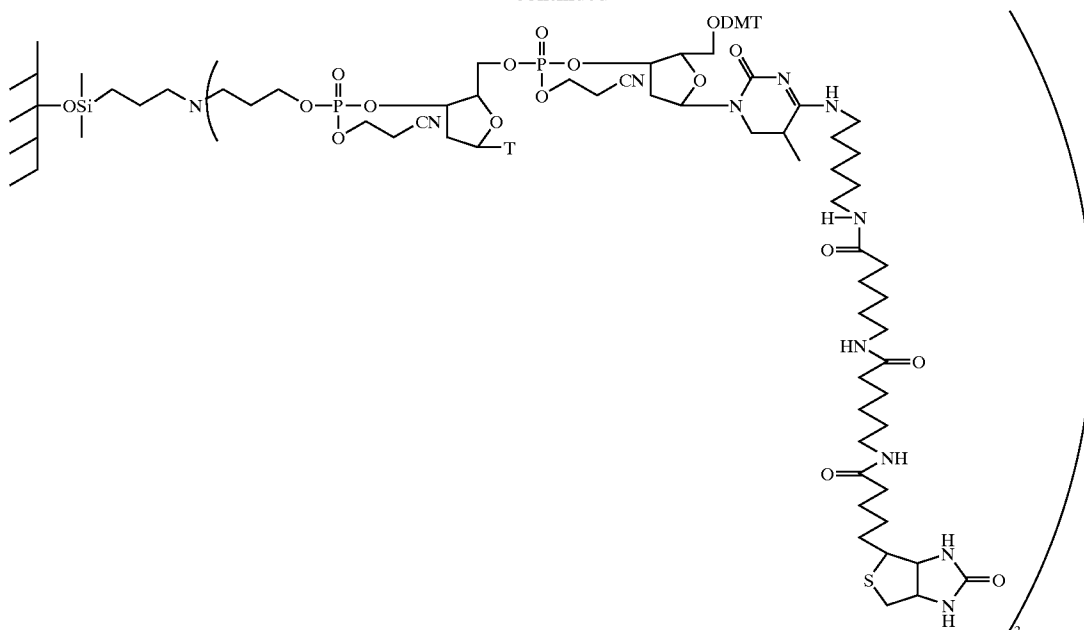


US 6,576,424 B2

65

66

-continued



3. Preparation and Detection of Thymidine-Cytidine Dimer (Biotin)

An aminopropyl glass slide, was soaked in a solution of ethylene oxide (20% in DMF) to generate a hydroxylated surface. The slide was added to a mixture of the following:

32 mg of NVO-T-OCED (X)

11 mg of tetrazole

0.5 ml of anhydrous CH_3CN

After 8 min the plate was then rinsed with acetonitrile, then oxidized with $\text{I}_2/\text{H}_2\text{O}/\text{THF}/\text{lutidine}$ for 1 min, washed and dried. The slide was then exposed to a 1:3 mixture of acetic anhydride:pyridine for 1 h, then washed and dried. The substrate was then photolyzed in dioxane at 362 nm at 14 mW/cm² for 10 min using a 500 μm checkerboard mask, dried, and then treated with a mixture of the following:

65 mg of biotin modified C (IV)

11 mg of tetrazole

0.5 ml anhydrous CH_3CN

After 8 min the slide was washed with CH_3CN then oxidized with $\text{I}_2/\text{H}_2\text{O}/\text{THF}/\text{lutidine}$ for 1 min, washed, and then dried. The slide was then soaked for 30 min in a PBS/0.05% Tween 20 buffer and the solution then shaken off. The slide was next treated with FITC-labeled streptavidin at 10 $\mu\text{g}/\text{ml}$ in the same buffer system for 30 min. After this time the streptavidin-buffer system was rinsed off with fresh PBS/0.05% Tween 20 buffer and then the slide was finally agitated in distilled water for about 1/2 h. After drying, the slide was examined by fluorescence microscopy.

4. Substrate Preparation

Before attachment of reactive groups it is preferred to clean the substrate which is, in a preferred embodiment, a glass substrate such as a microscope slide or cover slip. A roughened surface will be useable but a plastic or other solid substrate is also appropriate. According to one embodiment the slide is soaked in an alkaline bath consisting of, e.g., 1 liter of 95% ethanol with 120 ml of water and 120 grams of sodium hydroxide for 12 hours. The slides are washed with a buffer and under running water, allowed to air dry, and rinsed with a solution of 95% ethanol.

The slides are then aminated with, e.g., aminopropyltriethoxysilane for the purpose of attaching amino groups to the glass surface on linker molecules, although other omega functionalized silanes could also be used for this purpose. In one embodiment 0.1% aminopropyltriethoxysilane is utilized, although solutions with concentrations from 10⁻⁷% to 10% may be used, with about 10⁻³% to 2% preferred. A 0.1% mixture is prepared by adding to 100 ml of a 95% ethanol/5% water mixture, 100 microliters (μl) of aminopropyltriethoxysilane. The mixture is agitated at about ambient temperature on a rotary shaker for an appropriate amount of time, e.g., about 5 minutes. 500 μl of this mixture is then applied to the surface of one side of each cleaned slide. After 4 minutes or more, the slides are decanted of this solution and thoroughly rinsed three times or more by dipping in 100% ethanol.

After the slides dry, they are heated in a 110–120° C. vacuum oven for about 20 minutes, and then allowed to cure at room temperature for about 12 hours in an argon environment. The slides are then dipped into DMF (dimethylformamide) solution, followed by a thorough washing with methylene chloride.

5. Linker Attachment, Blocking of Free Sites

The aminated surface of the slide is then exposed to about 500 μl of, for example, a 30 millimolar (mM) solution of NVOC-nucleotide-NHS (N-hydroxysuccinimide) in DMF for attachment of a NVOC-nucleotide to each of the amino groups. See, e.g., SIGMA Chemical Company for various nucleotide derivatives. The surface is washed with, for example, DMF, methylene chloride, and ethanol.

Any unreacted aminopropyl silane on the surface, i.e., those amino groups which have not had the NVOC-nucleotide attached, are now capped with acetyl groups (to prevent further reaction) by exposure to a 1:3 mixture of acetic anhydride in pyridine for 1 hour. Other materials which may perform this residual capping function include trifluoroacetic anhydride, formicacetic anhydride, or other reactive acylating agents. Finally, the slides are washed again with DMF, methylene chloride, and ethanol.

US 6,576,424 B2

67

6. Synthesis of Eight Trimers of C and T

FIG. 2 illustrates a possible synthesis of the eight trimers of the two-monomer set: cytosine and thymine (represented by C and T, respectively). A glass slide bearing silane groups terminating in 6-nitroveratryloxycarbonyl (NVOC-NH) residues is prepared as a substrate. Active esters (pentafluorophenyl, OBT, etc.) of cytosine and thymine protected at the 5' hydroxyl group with NVOC are prepared as reagents. While not pertinent to this example, if side chain protecting groups are required for the monomer set, these must not be photoreactive at the wavelength of light used to protect the primary chain.

For a monomer set of size n , $n \times l$ cycles are required to synthesize all possible sequences of length l . A cycle consists of:

1. Irradiation through an appropriate mask to expose the 5'-OH groups at the sites where the next residue is to be added, with appropriate washes to remove the by-products of the deprotection.
2. Addition of a single activated and protected (with the same photochemically-removable group) monomer, which will react only at the sites addressed in step 1, with appropriate washes to remove the excess reagent from the surface.

The above cycle is repeated for each member of the monomer set until each location on the surface has been extended by one residue in one embodiment. In other embodiments, several residues are sequentially added at one location before moving on to the next location. Cycle times will generally be limited by the coupling reaction rate, now as short as about 10 min in automated oligonucleotide synthesizers. This step is optionally followed by addition of a protecting group to stabilize the array for later testing. For some types of polymers (e.g., peptides), a final deprotection of the entire surface (removal of photoprotective side chain groups) may be required.

More particularly, as shown in FIG. 2A, the glass 20 is provided with regions 22, 24, 26, 28, 30, 32, 34, and 36. Regions 30, 32, 34, and 36 are masked, indicated by the hatched regions, as shown in FIG. 2B and the glass is irradiated by the bright regions 22, 24, 26, and 28, and exposed to a reagent containing a photosensitive blocked C (e.g., cytosine derivative), with the resulting structure shown in FIG. 2C. The substrate is carefully washed and the reactants removed. Thereafter, regions 22, 24, 26, and 28 are masked, as indicated by the hatched region, the glass is irradiated (as shown in FIG. 2D), as indicated by the bright regions, at 30, 32, 34, and 36, and exposed to a photosensitive blocked reagent containing T (e.g., thymine derivative), with the resulting structure shown in FIG. 2E. The process proceeds, consecutively masking and exposing the sections as shown until the structure shown in FIG. 2M is obtained. The glass is irradiated and the terminal groups are, optionally, capped by acetylation. As shown, all possible trimers of cytosine/thymine are obtained.

In this example, no side chain protective group removal is necessary, as might be common in modified nucleotides. If it is desired, side chain deprotection may be accomplished by treatment with ethanedithiol and trifluoroacetic acid.

In general, the number of steps needed to obtain a particular polymer chain is defined by:

$$n \times l \quad (1)$$

where:

- n =the number of monomers in the basis set of monomers, and
 l =the number of monomer units in a polymer chain.

68

Conversely, the synthesized number of sequences of length l will be:

$$n^l. \quad (2)$$

Of course, greater diversity is obtained by using masking strategies which will also include the synthesis of polymers having a length of less than l . If, in the extreme case, all polymers having a length less than or equal to l are synthesized, the number of polymers synthesized will be:

$$n^l + n^{l-1} + \dots + n^1. \quad (3)$$

The maximum number of lithographic steps needed will generally be n for each "layer" of monomers, i.e., the total number of masks (and, therefore, the number of lithographic steps) needed will be $n \times l$. The size of the transparent mask regions will vary in accordance with the area of the substrate available for synthesis and the number of sequences to be formed. In general, the size of the synthesis areas will be:

$$\text{size of synthesis areas} = (A)/(S)$$

where:

A is the total area available for synthesis; and

S is the number of sequences desired in the area.

It will be appreciated by those of skill in the art that the above method could readily be used to simultaneously produce thousands or millions of oligomers on a substrate using the photolithographic techniques disclosed herein. Consequently, the method results in the ability to practically test large numbers of, for example, di, tri, tetra, penta, hexa, hepta, octa, nona, deca, even dodecanucleotides, or larger polynucleotides (or correspondingly, polypeptides).

The above example has illustrated the method by way of a manual example. It will of course be appreciated that automated or semi-automated methods could be used. The substrate would be mounted in a flow cell for automated addition and removal of reagents, to minimize the volume of reagents needed, and to more carefully control reaction conditions. Successive masks will be applicable manually or automatically. See, e.g., Pirrung et al. (1992) U.S. Pat. No. 5,143,854 and Ser. No. 07/624,120, now abandoned.

7. Labeling of Target

The target oligonucleotide can be labeled using standard procedures referred to above. As discussed, for certain situations, a reagent which recognizes interaction, e.g., ethidium bromide, may be provided in the detection step. Alternatively, fluorescence labeling techniques may be applied, see, e.g., Smith, et al. (1986) *Nature*, 321: 674-679; and Prober, et al. (1987) *Science*, 238:336-341. The techniques described therein will be followed with minimal modifications as appropriate for the label selected.

8. Dimers of A, C, G, and T

The described technique may be applied, with photosensitive blocked nucleotides corresponding to adenine, cytosine, guanine, and thymine, to make combinations of polynucleotides consisting of each of the four different nucleotides. All 16 possible dimers would be made using a minor modification of the described method.

9. 10-mers of A, C, G, and T

The described technique for making dimers of A, C, G, and T may be further extended to make longer oligonucleotides. The automated system described, e.g., in Pirrung et al. (1992) U.S. Pat. No. 5,143,854, and Ser. No. 07/624,120, now abandoned, can be adapted to make all possible 10-mers composed of the 4 nucleotides A, C, G, and T. The photosensitive, blocked nucleotide analogues have been

US 6,576,424 B2

69

described above, and would be readily adaptable to longer oligonucleotides.

10. Specific Recognition Hybridization to 10-mers

The described hybridization conditions are directly applicable to the sequence specific recognition reagents attached to the substrate, produced as described immediately above. The 10-mers have an inherent property of hybridizing to a complementary sequence. For optimum discrimination between full matching and some mismatch, the conditions of hybridization should be carefully selected, as described above. Careful control of the conditions, and titration of parameters should be performed to determine the optimum collective conditions.

11. Hybridization

Hybridization conditions are described in detail, e.g., in Hames and Higgins (1985) *Nucleic Acid Hybridisation: A Practical Approach*; and the considerations for selecting particular conditions are described, e.g., in Wetmur and Davidson, (1988) *J. Mol. Biol.* 31:349–370, and Wood et al. (1985) *Proc. Natl. Acad. Sci. USA* 82:1585–1588. As described above, conditions are desired which can distinguish matching along the entire length of the probe from where there is one or more mismatched bases. The length of incubation and conditions will be similar, in many respects, to the hybridization conditions used in Southern blot transfers. Typically, the GC bias may be minimized by the introduction of appropriate concentrations of the alkylammonium buffers, as described above.

Titration of the temperature and other parameters is desired to determine the optimum conditions for specificity and distinguishability of absolutely matched hybridization from mismatched hybridization.

A fluorescently labeled target or set of targets are generated, as described in Prober, et al. (1987) *Science* 238:336–341, or Smith, et al. (1986) *Nature* 321:674–679. Preferably, the target or targets are of the same length as, or slightly longer, than the oligonucleotide probes attached to the substrate and they will have known sequences. Thus, only a few of the probes hybridize perfectly with the target, and which particular ones did would be known.

The substrate and probes are incubated under appropriate conditions for a sufficient period of time to allow hybridization to completion. The time is measured to determine when the probe-target hybridizations have reached completion. A salt buffer which minimizes GC bias is preferred, incorporating, e.g., buffer, such as tetramethyl ammonium or tetraethyl ammonium ion at between about 2.4 and 3.0 M. See Wood, et al. (1985) *Proc. Nat'l Acad. Sci. USA* 82:1585–1588. This time is typically at least about 30 min., and may be as long as about 1–5 days. Typically very long matches will hybridize more quickly, very short matches will hybridize less quickly, depending upon relative target and probe concentrations. The hybridization will be performed under conditions where the reagents are stable for that time duration.

Upon maximal hybridization, the conditions for washing are titrated. Three parameters initially titrated are time, temperature, and cation concentration of the wash step. The matrix is scanned at various times to determine the conditions at which the distinguishability between true perfect hybrid and mismatched hybrid is optimized. These conditions will be preferred in the sequencing embodiments.

12. Positional Detection of Specific Interaction

As indicated above, the detection of specific interactions may be performed by detecting the positions where the labeled target sequences are attached. Where the label is a fluorescent label, the apparatus described, e.g., in Pirrung et

70

al. (1992) U.S. Pat. No. 5,143,854; and Ser. No. 07/624,120, now abandoned, may be advantageously applied. In particular, the synthetic processes described above will result in a matrix pattern of specific sequences attached to the substrate, and a known pattern of interactions can be converted to corresponding sequences.

In an alternative embodiment, a separate reagent which differentially interacts with the probe and interacted probe/targets can indicate where interaction occurs or does not occur. A single-strand specific reagent will indicate where no interaction has taken place, while a double-strand specific reagent will indicate where interaction has taken place. An intercalating dye, e.g., ethidium bromide, may be used to indicate the positions of specific interaction.

13. Analysis

Conversion of the positional data into sequence specificity will provide the set of subsequences whose analysis by overlap segments, may be performed, as described above. Analysis is provided by the methodology described above, or using, e.g., software available from the Genetic Engineering Center, P.O. Box 794, 11000 Belgrade, Yugoslavia (Yugoslav group). See, also, Macevitz, PCT publication no. WO 90/04652, which is hereby incorporated herein by reference.

B. Polypeptide

The description of the preparation of short peptides on a substrate incorporates by reference sections in Pirrung et al. (1992) U.S. Pat. No. 5,143,854, and described below.

1. Slide Preparation

Preparation of the substrate follows that described above for nucleotides.

2. Linker Attachment, Blocking of Free Sites

The aminated surface of the slide is exposed to about 500 μ l of, e.g., a 30 millimolar (mM) solution of NVOC-GABA (gamma amino butyric acid) NHS (N-hydroxysuccinimide) in DMF for attachment of a NVOC-GABA to each of the amino groups. The surface is washed with, for example, DMF, methylene chloride, and ethanol. See Ser. No. 07,624, 120, now abandoned, for details on amino acid chemistry.

Any unreacted aminopropyl silane on the surface, i.e., those amino groups which have not had the NVOC-GABA attached, are now capped with acetyl groups (to prevent further reaction) by exposure to a 1:3 mixture of acetic anhydride in pyridine for 1 hour. Other materials which may perform this residual capping function include trifluoroacetic anhydride, formicacetic anhydride, or other reactive acylating agents. Finally, the slides are washed again with DMF, methylene chloride, and ethanol.

3. Synthesis of 8 Trimers of "A" and "B"

See Pirrung et al. (1992) U.S. Pat. No. 5,143,854 which describes the preparation of glycine and phenylalanine trimers. The technique is similar to the method described above for making trimers of C and T, but substituting photosensitive blocked glycine for the C derivative and photosensitive blocked phenylalanine for the T derivative.

4. Synthesis of a Dimer of an Aminopropyl Group and a Fluorescent Group

In synthesizing the dimer of an aminopropyl group and a fluorescent group, a functionalized Durapore™ membrane was used as a substrate. The Durapore™ membrane was a polyvinylidene difluoride with aminopropyl groups. The aminopropyl groups were protected with the DDZ group by reaction of the carbonyl chloride with the amino, groups, a reaction readily known to those of skill in the art. The surface bearing these groups was placed in a solution of THF and contacted with a mask bearing a checkerboard pattern of 1 mm opaque and transparent regions. The mask was

US 6,576,424 B2

71

exposed to ultraviolet light having a wavelength down to at least about 280 nm for about 5 minutes at ambient temperature, although a wide range of exposure times and temperatures may be appropriate in various embodiments of the invention. For example, in one embodiment, an exposure time of between about 1 and 5000 seconds may be used at process temperatures of between -70 and $+50^{\circ}\text{C}$.

In one preferred embodiment, exposure times of between about 1 and 500 seconds at about ambient pressure are used. In some preferred embodiments, pressure above ambient is used to prevent evaporation.

The surface of the membrane was then washed for about 1 hour with a fluorescent label which included an active ester bound to a chelate of a lanthanide. Wash times will vary over a wide range of values from about a few minutes to a few hours. These materials fluoresce in the red and the green visible region. After the reaction with the active ester in the fluorophore was complete, the locations in which the fluorophore was bound could be visualized by exposing them to ultraviolet light and observing the red and the green fluorescence. It was observed that the derivatized regions of the substrate closely corresponded to the original pattern of the mask.

5. Demonstration of Signal Capability

Signal detection capability was demonstrated using a low-level standard fluorescent bead kit manufactured by Flow Cytometry Standards and having model no. 824. This kit includes $5.8\text{ }\mu\text{m}$ diameter beads, each impregnated with a known number of fluorescein molecules.

One of the beads was placed in the illumination field on the scan stage in a field of a laser spot which was initially shuttered. After being positioned in the illumination field, the photon detection equipment was turned on. The laser beam was unblocked and it interacted with the particle bead, which then fluoresced. Fluorescence curves of beads impregnated with 7,000 and 29,000 fluorescein molecules, are shown in FIGS. 11A and 11B, respectively of Pirrung et al. (1992) U.S. Pat. No. 5,143,854. On each curve, traces for beads without fluorescein molecules are also shown. These experiments were performed with 488 nm excitation, with 100 μW of laser power. The light was focused through a 40 power 0.75 NA objective.

The fluorescence intensity in all cases started off at a high value and then decreased exponentially. The fall-off in intensity is due to photobleaching of the fluorescein molecules. The traces of beads without fluorescein molecules are used for background subtraction. The difference in the initial exponential decay between labeled and nonlabeled beads is integrated to give the total number of photon counts, and this number is related to the number of molecules per bead. Therefore, it is possible to deduce the number of photons per fluorescein molecule that can be detected. This calculation indicates the radiation of about 40 to 50 photons per fluorescein molecule are detected.

6. Determination of the Number of Molecules Per Unit Area

Aminopropylated glass microscope slides prepared according to the methods discussed above were utilized in order to establish the density of labeling of the slides. The free amino termini of the slides were reacted with FITC (fluorescein isothiocyanate) which forms a covalent linkage with the amino group. The slide is then scanned to count the number of fluorescent photons generated in a region which, using the estimated 40–50 photons per fluorescent molecule, enables the calculation of the number of molecules which are on the surface per unit area.

A slide with aminopropyl silane on its surface was immersed in a 1 mM solution of FITC in DMF for 1 hour at

72

about ambient temperature. After reaction, the slide was washed twice with DMF and then washed with ethanol, water, and then ethanol again. It was then dried and stored in the dark until it was ready to be examined.

Through the use of curves similar to those shown in FIG. 11 of Pirrung et al. (1992) U.S. Pat. No. 5,143,854, and by integrating the fluorescent counts under the exponentially decaying signal, the number of free amino groups on the surface after derivitization was determined. It was determined that slides with labeling densities of 1 fluorescein per $10^3 \times 10^3$ to $\sim 2 \times 2$ nm could be reproducibly made as the concentration of aminopropyltriethoxysilane varied from $10^{-5}\%$ to $10^{-1}\%$.

7. Removal of NVOC and Attachment of a Fluorescent Marker

NVOC-GABA groups were attached as described above. The entire surface of one slide was exposed to light so as to expose a free amino group at the end of the gamma amino butyric acid. This slide, and a duplicate which was not exposed, were then exposed to fluorescein isothiocyanate (FITC).

FIG. 12A of Pirrung et al. (1992) U.S. Pat. No. 5,143,854 illustrates the slide which was not exposed to light, but which was exposed to FIT

C. The units of the x axis are time and the units of the y axis are counts. The trace contains a certain amount of background fluorescence. The duplicate slide was exposed to 350 nm broadband illumination for about 1 minute (12 mW/cm², ~ 350 nm illumination), washed and reacted with FIT

C. A large increase in the level of fluorescence is observed, which indicates photolysis has exposed a number of amino groups on the surface of the slides for attachment of a fluorescent marker.

8. Use of a Mask in Removal of NVOC

The next experiment was performed with a 0.1% aminopropylated slide. Light from a Hg-Xe arc lamp was imaged onto the substrate through a laser-ablated chrome-on-glass mask in direct contact with the substrate.

This slide was illuminated for approximately 5 minutes, with 12 mW of 350 nm broadband light and then reacted with the 1 mM FITC solution. It was put on the laser detection scanning stage and a graph was plotted as a two-dimensional representation of position color-coded for fluorescence intensity. The experiment was repeated a number of times through various masks. The fluorescence patterns for a $100 \times 100\text{ }\mu\text{m}$ mask, a $50\text{ }\mu\text{m}$ mask, a $20\text{ }\mu\text{m}$ mask, and a $10\text{ }\mu\text{m}$ mask indicate that the mask pattern is distinct down to at least about $10\text{ }\mu\text{m}$ squares using this lithographic technique.

9. Attachment of YGGFL and Subsequent Exposure to Herz Antibody and Goat Anti-mouse Antibody

In order to establish that receptors to a particular polypeptide sequence would bind to a surface-bound peptide and be detected, Leu enkephalin was coupled to the surface and recognized by an antibody. A slide was derivatized with 0.1% amino propyl-triethoxysilane and protected with NVOC. A $500\text{ }\mu\text{m}$ checkerboard mask was used to expose the slide in a flow cell using backside contact printing. The Leu enkephalin sequence (H_2N -tyrosine, glycine, glycine, phenylalanine, leucine-COOH, otherwise referred to herein as YGGFL) was attached via its carboxy end to the exposed amino groups on the surface of the slide. The peptide was added in DMF solution with the BOP/HOBT/DIEA coupling reagents and recirculated through the flow cell for 2 hours at room temperature.

A first antibody, known as the Herz antibody, was applied to the surface of the slide for 45 minutes at $2\text{ }\mu\text{g/ml}$ in a

supercocktail (containing 1% BSA and 1% ovalbumin also in this case). A second antibody, goat anti-mouse fluorescein conjugate, was then added at 2 μ g/ml in the supercocktail buffer, and allowed to incubate for 2 hours.

The results of this experiment were plotted as fluorescence intensity as a function of position. This image was taken at 10 μ m steps and showed that not only can deprotection be carried out in a well defined pattern, but also that (1) the method provided for successful coupling of peptides to the surface of the substrate, (2) the surface of a bound peptide was available for binding with an antibody, and (3) the detection apparatus capabilities were sufficient to detect binding of a receptor. Moreover, the Herz antibody is a sequence specific reagent which may be used advantageously as a sequence specific recognition reagent. It may be used, if specificity is high, for sequencing purposes, and, at least, for fingerprinting and mapping uses.

10. Monomer-by-Monomer Formation of YGGFL and Subsequent Exposure to Labeled Antibody

Monomer-by-monomer synthesis of YGGFL and GGFL in alternate squares was performed on a slide in a checkerboard pattern and the resulting slide was exposed to the Herz antibody.

A slide is derivatized with the aminopropyl group, protected in this case with t-BOC (t-butoxycarbonyl). The slide was treated with TFA to remove the t-BOC protecting group. E-aminocaproic acid, which was t-BOC protected at its amino group, was then coupled onto the aminopropyl groups. The aminocaproic acid serves as a spacer between the aminopropyl group and the peptide to be synthesized. The amino end of the spacer was deprotected and coupled to NVOC-leucine. The entire slide was then illuminated with 12 mW of 325 nm broadband illumination. The slide was then coupled with NVOC-phenylalanine and washed. The entire slide was again illuminated, then coupled to NVOC-glycine and washed. The slide was again illuminated and coupled to NVOC-glycine to form the sequence shown in the last portion of FIG. 13A of Pirrung et al. (1992) U.S. Pat. No. 5,143,854.

Alternating regions of the slide were then illuminated using a projection print using a 500x500 μ m checkerboard mask; thus, the amino group of glycine was exposed only in the lighted areas. When the next coupling chemistry step was carried out, NVOC-tyrosine was added, and it coupled only at those spots which had received illumination. The entire slide was then illuminated to remove all the NVOC groups, leaving a checkerboard of YGGFL in the lighted areas and in the other areas, GGFL. The Herz antibody (which recognizes the YGGFL, but not GGFL) was then added, followed by goat anti-mouse fluorescein conjugate.

The resulting fluorescence scan showed dark areas containing the tetrapeptide GGFL, which is not recognized by the Herz antibody (and thus there is no binding of the goat anti-mouse antibody with fluorescein conjugate), and red areas in which YGGFL was present. The YGGFL pentapeptide is recognized by the Herz antibody and, therefore, there is antibody in the lighted regions for the fluorescein-conjugated goat anti-mouse to recognize.

Similar patterns for a 50 μ m mask used in direct contact ("proximity print") with the substrate provided a pattern which was more distinct and the corners of the checkerboard pattern were touching as a result of the mask being placed in direct contact with the substrate (which reflects the increase in resolution using this technique).

11. Monomer-by-monomer Synthesis of YGGFL and PGGFL

A synthesis using a 50 μ m checkerboard mask was conducted. However, P was added to the GGFL sites on the

substrate through an additional coupling step. P was added by exposing protected GGFL to light through a mask, and subsequent exposure to P in the manner set forth above. Therefore, half of the regions on the substrate contained YGGFL and the remaining half contained PGGFL.

The fluorescence plot for this experiment showed the regions are again readily discernable between those in which binding did and did not occur. This experiment demonstrated that antibodies are able to recognize a specific sequence and that the recognition is not length-dependent.

12. Monomer-by-Monomer Synthesis of YGGFL and YPGGFL

In order to further demonstrate the operability of the invention, a 50 μ m checkerboard pattern of alternating YGGFL and YPGGFL was synthesized on a substrate using techniques like those set forth above. The resulting fluorescence plot showed that the antibody was clearly able to recognize the YGGFL sequence and did not bind significantly at the YPGGFL regions.

13. Synthesis of an Array of Sixteen Different Amino Acid Sequences and Estimation of Relative Binding Affinity to Herz Antibody

Using techniques similar to those set forth above, an array of 16 different amino acid sequences (replicated four times) was synthesized on each of two glass substrates. The sequences were synthesized by attaching the sequence NVOC-GFL across the entire surface of the slides. Using a series of masks, two layers of amino acids were then selectively applied to the substrate. Each region had dimensions of 0.25 cmx0.0625 cm. The first slide contained amino acid sequences containing only L-amino acids while the second slide contained selected D-amino acids. Various regions on the first and second slides, were duplicated four times on each slide. The slides were then exposed to the Herz antibody and fluorescein-labeled goat anti-mouse antibodies.

A fluorescence plot of the first slide, which contained only L-amino acids showed red areas (indicating strong binding, i.e., 149,000 counts or more) and black areas (indicating little or no binding of the Herz antibody, i.e., 20,000 counts or less). The sequence YGGFL was clearly most strongly recognized. The sequences YAGFL and YSGFL also exhibited strong recognition of the antibody. By contrast, most of the remaining sequences showed little or no binding. The four duplicate portions of the slide were extremely consistent in the amount of binding shown therein.

A fluorescence plot of the D-amino acid slide indicated that strongest binding was exhibited by the YGGFL sequence. Significant binding was also detected to YaGFL, YsGFL, and YpGFL. The remaining sequences showed less binding with the antibody. Low binding efficiency of the sequence yGGFL was observed.

Table 6 lists the various sequences tested in order of relative fluorescence, which provides information regarding relative binding affinity.

TABLE 6

Apparent Binding to Herz Ab	
L- a.a. Set	D- a.a. Set
YGGFL	YGGFL
YAGFL	YaGFL
YSGL	YsGFL
LGGFL	YpGFL
FGGFL	fGGFL
YPGFL	yGGFL

TABLE 6-continued

<u>Apparent Binding to Herz Ab</u>	
L- a.a. Set	D- a.a. Set
LAGFL	faGFL
FAGFL	wGGFL
WGGFL	yaGFL
	fpGFL
	waGFL

14. Illustrative Alternative Embodiment

According to an alternative embodiment of the invention, the methods provide for attaching to the surface a caged binding member which, in its caged form, has a relatively low affinity for other potentially binding species, such as receptors and specific binding substances. Such techniques are more fully described in copending application Ser. No. 404,920, filed Sep. 8, 1989, and incorporated herein by reference for all purposes. See also Ser. No. 07/435,316, now abandoned, and Barrett et al. (1993) U.S. Pat. No. 5,252,743, each of which is hereby incorporated herein by reference.

According to this alternative embodiment, the invention provides methods for forming predefined regions on a surface of a solid support, wherein the predefined regions are capable of immobilizing receptors. The methods make use of caged binding members attached to the surface to enable selective activation of the predefined regions. The caged binding members are liberated to act as binding members ultimately capable of binding receptors upon selective activation of the predefined regions. The activated binding members are then used to immobilize specific molecules such as receptors on the predefined region of the surface. The above procedure is repeated at the same or different sites on the surface so as to provide a surface prepared with a plurality of regions on the surface containing, for example, the same or different receptors. When receptors immobilized in this way have a differential affinity for one or more ligands, screenings and assays for the ligands can be conducted in the regions of the surface containing the receptors.

The alternative embodiment may make use of novel caged binding members attached to the substrate. Caged (unactivated) members have a relatively low affinity for receptors of substances that specifically bind to uncaged binding members when compared with the corresponding affinities of activated binding members. Thus, the binding members are protected from reaction until a suitable source of energy is applied to the regions of the surface desired to be activated. Upon application of a suitable energy source, the caging groups labilize, thereby presenting the activated binding member. A typical energy source will be light.

Once the binding members on the surface are activated they may be attached to a receptor. The receptor chosen may be a monoclonal antibody, a nucleic acid sequence, a drug receptor, etc. The receptor will usually, though not always, be prepared so as to permit attaching it, directly or indirectly, to a binding member. For example, a specific binding substance having a strong binding affinity for the binding member and a strong affinity for the receptor or a conjugate of the receptor may be used to act as a bridge between binding members and receptors if desired. The method uses a receptor prepared such that the receptor retains its activity toward a particular ligand.

Preferably, the caged binding member attached to the solid substrate will be a photoactivatable biotin complex, i.e., a biotin molecule that has been chemically modified

with photoactivatable protecting groups so that it has a significantly reduced binding affinity for avidin or avidin analogs than does natural biotin. In a preferred embodiment, the protecting groups localized in a predefined region of the surface will be removed upon application of a suitable source of radiation to give binding members, that is biotin or a functionally analogous compound having substantially the same binding affinity for avidin or avidin analogs as does biotin.

In another preferred embodiment, avidin or an avidin analog is incubated with activated binding members on the surface until the avidin binds strongly to the binding members. The avidin so immobilized on predefined regions of the surface can then be incubated with a desired receptor or conjugate of a desired receptor. The receptor will preferably be biotinylated, e.g., a biotinylated antibody, when avidin is immobilized on the predefined regions of the surface. Alternatively, a preferred embodiment will present an avidin/biotinylated receptor complex, which has been previously prepared, to activated binding members on the surface.

II. FINGERPRINTING

The above section on generation of reagents for sequencing provides specific reagents useful for fingerprinting applications. Fingerprinting embodiments may be applied towards polynucleotide fingerprinting, polypeptide fingerprinting, cell and tissue classification, cell and tissue temporal development stage classification, diagnostic tests, forensic uses for individual identification, classification of organisms, and genetic screening of individuals. Mapping applications are also described below.

A. Polynucleotide Fingerprint

Polynucleotide fingerprinting may use reagents similar to those described above for probing a sequence for the presence of specific subsequences found therein. Typically, the subsequences used for fingerprinting will be longer than the sequences used in oligonucleotide sequencing. In particular, specific long segments may be used to determine the similarity of different samples of nucleic acids. They may also be used to fingerprint whether specific combinations of information are provided therein. Particular probe sequences are selected and attached in a positional manner to a substrate. The means for attachment may be either using a caged biotin method described, e.g., in Barrett et al. (1993) U.S. Pat. No. 5,242,743, or by another method using targeting molecules. For example, a short polypeptide of specific sequence may be attached to an oligonucleotide and targeted to specific positions on a substrate having antibodies attached thereto, the antibodies exhibiting specificity for binding to those short peptide sequences. In another embodiment, an unnatural nucleotide or similar complementary binding molecule may be attached to the fingerprinting probe and the probe thereby directed towards complementary sequences on a VLSIPS substrate. Typically, unnatural nucleotides would be preferred, e.g., unnatural optical isomers, which would not interfere with natural nucleotide interactions.

Having produced a substrate with particular fingerprint probes attached thereto at positionally defined regions, the substrate may be used in a manner quite similar to the sequencing embodiment to provide information as to whether the fingerprint probes are detecting the corresponding sequence in a target sequence. This will often provide information similar to a Southern blot hybridization.

B. Polypeptide Fingerprint

A polypeptide fingerprint may be performed using antibodies which recognize specific antigens on the polypeptide.

US 6,576,424 B2

77

For example, monoclonal antibodies which recognize specific sequences or antigens on a polypeptide may be used to determine whether those epitopes are found on a particular protein. For example, particular patterns of epitopes would be found on various types of proteins. This will lead to the discovery that specific epitopes, or antigenic determinants, which are characteristic of, e.g., beta sheet segments, will be identified as will particular different types of domains in various protein types. Thus, a screening method may be devised which can classify polypeptides, either native or denatured, into various new classes defined by the epitopes existing thereon.

In addition, once the substrate is generated in the manners described above, a target peptide is exposed to the substrate. The target may be either native or denatured, though the conditions used to denature the polypeptide may interfere with the specific interaction between the polypeptide and the recognition reagent. This method is not dependent on the fact that the polypeptide is a single chain, thus protein complexes may also be fingerprinted using this methodology. Structures such as multi-subunit proteins, associations of proteins, ribosomes, nucleosomes, and other small cellular structures may also be fingerprinted and classified according to the presence of specific recognizable features thereon.

Peptide fingerprinting may be useful, for example, in correlating with particular physiological conditions or developmental stages of a cell or organism. Thus, a biological sample may be fingerprinted to determine the presence in that sample of a plurality of different polypeptides which are each individually fingerprinted. In an alternative embodiment, a polypeptide itself is not fingerprinted but a biological sample is fingerprinted searching for specific epitopes, e.g., polypeptide, carbohydrate, nucleic acid, or any of a number of other specific recognizable structural features.

The conditions for the interactions using antibodies is described, e.g., in Harlow and Lane (1988) *Antibodies: A Laboratory Manual*, Cold Spring Harbor Press, New York. The conditions should be titrated for temperature, buffer composition, time, and other important parameters in an antibody interaction.

C. Cell Classification Scheme

The present invention can be used for cell classification using fingerprinting type technology as described above in the polypeptide fingerprint. Classes of cells are typically defined by the presence of common functions which are usually reflected by structural features. Thus, a plant cell is classified differently from an animal cell by a number of structural features. Given an unknown cell, the present invention provides improved means for distinguishing the different cell types. Once a cell classification scheme is developed and the structural features which define it are identified using the present invention, homogeneous cell population expressing these features may be separated from others. Standard cell sorters may be coupled with recognition reagents and labels which can distinguish various cell types.

a. T-Cell Classes

T-cell classes are defined on the basis of expression of particular antigens characteristic of each class. For example, mouse T-cell differentiation markers include the LY antigens. With the plurality of different antigens which may be tested using antibody or other recognition reagents, new populations and classes of cells may be defined. For example, different neural cell types may be defined on the basis of cell surface antigens. Different tissue types will be

78

defined on the basis of tissue specific antigens. Developmental cell classes will be similarly defined. All of these screenings can make use of the VLSIPS substrates with specific recognition molecules attached thereto. The substrates are exposed to the cell types directly, assaying for attachment of cells to specific regions, or are exposed to products of a population of cells, e.g., a supernatant, or a cell lysate.

Once a cell classification scheme has been correlated with specific structural markers therein, reagents which recognize those features may be developed and used in a fluorescence activated cell sorter as described, e.g., in Dangi, J. and Herzenberg (1982) *J. Immunological Methods* 52: 1-14; and Becton Dickinson, Fluorescence Activated Cell Sorters Division, San Jose, Calif. This will provide a homogeneous population of cells whose function has been defined by structure.

b. B-Cell Classes

The present cell classification scheme may also be used to determine specific B-cell classes. For example, B-cells specific for producing IgM, IgG, IgD, IgE, and IgA may be defined by the internal expression of specific mRNA sequences encoding each type of immunoglobulin. The classification scheme may depend on either extracellularly expressed markers which are correlated as being diagnostic of specific stages in development, or intracellular mRNA sequences which indicate particular functions.

D. Temporal Development Scheme

1. Developmental Antigens

The present fingerprinting invention also allows cell classification by expression of developmental antigens. For example, a lymphocyte stem cell expresses a particular combination of antigens. As the lymphocyte develops through a program developmental scheme, at various stages it expresses particular antigens which are diagnostic of particular stages in development. Again, the fingerprinting methodology allows for the definition of specific structural features which are diagnostic of developmental or functional features which will allow classification of cells into temporal developmental classes. Cells, products of those cells, or lysates of those cells will be assayed to determine the developmental stage of the source cells. In this manner, once a developmental stage is defined, specific synchronized populations of cells will be selected out of another population. These synchronized populations may be very important in determining the biological mechanisms of development.

2. Developmental mRNA Expression

Besides expressed antigens, the present invention also allows for fingerprinting of the mRNA population of a cell. In this fashion, the mRNA population, which should be a good determinant of developmental stage, will be correlated with other structural features of the cell. In this manner, cells at specific developmental stages will be characterized by the intracellular environment, as well as the extracellular environment. The present invention also allows the combination of definitions based, in part, upon antigens and, in part, upon mRNA expression.

In one embodiment, the two may be combined in a single incubation step. A particular incubation condition may be found which is compatible with both hybridization recognition and non-hybridization recognition molecules. Thus, e.g., an incubation condition may be selected which allows both specificity of antibody binding and specificity of nucleic acid hybridization. This allows simultaneous performance of both types of interactions on a single matrix. Again, where developmental mRNA patterns are correlated with structural features, or with probes which are able to

US 6,576,424 B2

79

hybridize to intracellular mRNA populations, a cell sorter may be used to sort specifically those cells having desired mRNA population patterns.

E. Diagnostic Tests

The present invention also provides the ability to perform diagnostic tests. Diagnostic tests typically are based upon a fingerprint type assay, which tests for the presence of specific diagnostic structural features. Thus, the present invention provides means for viral strain identification, bacterial strain identification, and other diagnostic tests using positionally defined specific reagents. The present invention also allows for determining a spectrum of allergies, diagnosing a biological sample for any or all of the above, and testing for many other conditions.

1. Viral Identification

The present invention provides reagents and methodology for identifying viral strains. The specific reagents may be either antibodies or recognition proteins which bind to specific viral epitopes preferably surface exposed, but may make use of internal epitopes, e.g., in a denatured viral sample. In an alternative embodiment, the viral genome may be probed for specific sequences which are characteristic of particular viral strains. As above, a combination of the two may be performed simultaneously in a single interaction step, or in separate tests, e.g., for both genetic characteristics and epitope characteristics.

2. Bacterial Identification

Similar techniques will be applicable to identifying a bacterial source. This may be useful in diagnosing bacterial infections, or in classifying sources of particular bacterial species. For example, the bacterial assay may be useful in determining the natural range of survivability of particular strains of bacteria across regions of the country or in different ecological niches.

3. Other Microbiological Identifications

The present invention provides means for diagnosis of other microbiological and other species, e.g., protozoal species and parasitic species in a biological sample, but also provides the means for assaying a combination of different infections. For example, a biological specimen may be assayed for the presence of any or all of these microbiological species. In human diagnostic uses, typical samples will be blood, sputum, stool, urine, or other samples.

4. Allergy Tests

An immobilized set of antigens may be attached to a solid substrate and, instead of the standard skin reaction tests, a blood sample may be assayed on such a substrate to determine the presence of antibodies, e.g., IgE or other type antibodies, which may be diagnostic of an allergic or immunological susceptibility. A standard radioallergosorbent test (RAST) may be used to check a much larger population of antigens.

In addition, an allergy like test may be used to diagnose the immunological history of a particular individual. For example, by testing the circulating antibodies in a blood sample, which reflects the immunological history and memory of an individual, it may be determined what infections may not have been historically presented to the immune system. In this manner, it may be possible to specifically supplement an immune system for a short period of time with IgG fractions made up of specific types of gamma globulins. Thus, hepatitis gamma globulin injections may be better designed for a particular environment to which a person is expected to be exposed. This also provides the ability to identify genetically equivalent individuals who have immunologically different experiences. Thus, a blood sample from an individual who has a particular combination

80

of circulating antibodies will likely be different from the combination of circulating antibodies found in a genetically similar or identical individual. This could allow for the distinction between clones of particular animals, e.g., mice, rats, or other animals.

F. Individual Identification

The present invention provides the ability to fingerprint and identify a genetic individual. This individual may be a bacterial or lower microorganism, as described above in diagnostic tests, or of a plant or animal. An individual may be identified genetically or immunologically, as described.

1. Genetic

Genetic fingerprinting has been utilized in comparing different related species in Southern hybridization blots. Genetic fingerprinting has also been used in forensic studies, see, e.g., Morris et al. (1989) *J. Forensic Science* 34: 1311-1317, and references cited therein. As described above, an individual may be identified genetically by a sufficiently large number of probes. The likelihood that another individual would have an identical pattern over a sufficiently large number of probes may be statistically negligible. However, it is often quite important that a large number of probes be used where the statistical probability of matching is desired to be particularly low. In fact, the probes will optimally be selected for having high heterogeneity among the population. In addition, the fingerprint method may make use of the pattern of homologies indicated by a series of more and more stringent washes. Then, each position has both a sequence specificity and a homology measurement, the combination of which greatly increases the number of dimensions and the statistical likelihood of a perfect pattern match with another genetic individual.

2. Immunological

As indicated above in the diagnostic tests, it is possible to identify a particular immune system within a genetically homogeneous class of organisms by virtue of their immunological history. For example, a large colony of cloned mice may be distinguishable by virtue of each immunological history. For example, one mouse may have had an immunological response to exposure to antigen A to which her genetically identical sibling may have not been exposed. By virtue of this differential history, the first of the pair will likely have a high antibody titer against the antigen A whereas her genetically identical sibling will have not had a response to that antigen by virtue of never having been exposed to it. For this reason, immune systems may be identified by their immunological memories. Thus, immunological experience may also be a means for identifying a particular individual at a particular moment in her lifetime.

This same immunological screening may be used for other sorts of identifiable biological products. For example, an individual may be identified by her combination of expressed proteins. These proteins may reflect a physiological state of the individual, and would thus be useful in certain circumstances where diagnostic tests may be performed. For example, an individual may be identified, in part, by the presence of particular metabolic products.

In fact, a plant origin may be determined by virtue of having within its genome an unnatural sequence introduced to it by genetic breeders. Thus, a marker nucleic acid sequence may be introduced as a means to determine whether a genetic strain of a plant or animal originated from another particular source.

G. Genetic Screening

1. Test Alleles with Markers

The present invention provides for the ability to screen for genetic variations of individuals. For example, a number of

US 6,576,424 B2

81

genetic diseases are linked with specific alleles. See, e.g., Scriber, C. et al. (eds.) (1989) *The Metabolic Bases of Inherited Disease*, McGraw-Hill, New York. In one embodiment, cystic fibrosis has been correlated with a specific gene, see, Gregory et al. (1990) *Nature* 347: 382–386. A number of alleles are correlated with specific genetic deficiencies. See, e.g., McKusick, V. (1990) *Genetic Inheritance in Man: Catalogs of Autosomal Dominant, Autosomal Recessive, and X-linked Phenotypes*, Johns Hopkins University Press, Baltimore; Ott, J. (1985) *Analysis of Human Genetic Linkage*, Johns Hopkins University Press, Baltimore; Track, R. et al. (1989) *Banbury Report 32: DNA Technology and Forensic Science*, Cold Spring Harbor Press, New York; each of which is hereby incorporated herein by reference.

2. Amniocentesis

Typically, amniocentesis is used to determine whether chromosome translocations have occurred. The mapping procedure may provide the means for determining whether these translocations have occurred, and for detecting particular alleles of various markers.

III. MAPPING

A. Positionally Located Clones

The present invention allows for the positional location of specific clones useful for mapping. For example, caged biotin may be used for specifically positioning a probe to a location on a matrix pattern.

In addition, the specific probes may be positionally directed to specific locations on a substrate by targeting. For example, polypeptide specific recognition reagents may be attached to oligonucleotide sequences which can be complementarily targeted to specific locations on a VLSIPS™ Technology substrate. Hybridization conditions, as applied for oligonucleotide probes, will be used to target the reagents to locations on a substrate having complementary oligonucleotides synthesized thereon. In another embodiment, oligonucleotide probes may be attached to specific polypeptide targeting reagents such as an antigen or antibody. These reagents can be directed towards a complementary antigen or antibody already attached to a VLSIPS substrate.

In another embodiment, an unnatural nucleotide which does not interfere with natural nucleotide complementary hybridization may be used to target oligonucleotides to particular positions on a substrate. Unnatural optical isomers of natural nucleotides should be ideal candidates.

In this way, short probes may be used to determine the mapping of long targets or long targets may be used to map the position of shorter probes. See, e.g., Craig et al. 1990 *Nuc. Acids Res.* 18: 2653–2660.

B. Positionally Defined Clones

Positionally defined clones may be transferred to a new substrate by either physical transfer or by synthetic means. Synthetic means may involve either a production of the probe on the substrate using the VLSIPS™ Technology synthetic methods, or may involve the attachment of a targeting sequence made by VLSIPS synthetic methods which will target that positionally defined clone to a position on a new substrate. Both methods will provide a substrate having a number of positionally defined probes useful in mapping.

IX. CONCLUSION

The present inventions provide greatly improved methods and apparatus for synthesis of polymers on substrates. It is to be understood that the above description is intended to be

82

illustrative and not restrictive. Many embodiments will be apparent to those of skill in the art upon reviewing the above description. By way of example, the invention has been described primarily with reference to the use of photoremovable protective groups, but it will be readily recognized by those of skill in the art that sources of radiation other than light could also be used. For example, in some embodiments it may be desirable to use protective groups which are sensitive to electron beam irradiation, x-ray irradiation, in combination with electron beam lithograph, or x-ray lithography techniques. Alternatively, the group could be removed by exposure to an electric current. The scope of the invention should, therefore, be determined not with reference to the above description, but should instead be determined with reference to the appended claims, along with the full scope of equivalents to which such claims are entitled.

All publications and patent applications referred to herein are incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually incorporated by reference. The present invention now being fully described, it will be apparent to one of ordinary skill in the art that many changes and modifications can be made thereto without departing from the spirit or scope of the appended claims.

What is claimed is:

1. An array comprising a solid substrate and a plurality of positionally distinguishable sequence specific polynucleotides attached to the solid substrate of at least 100 polynucleotides per cm², at least a plurality of said polynucleotides comprising at least 25 nucleotides.

2. The array of claim 1, comprising at least 300 of said positionally distinguishable polynucleotides per cm² attached to the solid substrate.

3. The array of claim 1, comprising at least 1000 of said positionally distinguishable polynucleotides per cm² attached to the solid substrate.

4. The array of claim 1, comprising at least 3000 of said positionally distinguishable polynucleotides per cm² attached to the solid substrate.

5. The array of claim 1, comprising at least 10,000 of said positionally distinguishable polynucleotides per cm² attached to the solid substrate.

6. The array of claim 1, comprising at least 30,000 of said positionally distinguishable polynucleotides per cm² attached to the solid substrate.

7. The array of claim 1, comprising at least 100,000 of said positionally distinguishable polynucleotides per cm² attached to the solid substrate.

8. The array of claim 1, wherein said plurality of attached polynucleotides comprises at least 3000 different polynucleotides.

9. The array of claim 1, wherein the solid substrate has a surface area of less than 4 square centimeters.

10. The array of claim 1, wherein the solid substrate has a surface area of less than 11 cm².

11. The array of claim 1, wherein the solid substrate is glass.

12. The array of claim 1, wherein the solid substrate comprises a polymeric substrate.

13. The array of claim 1, wherein the solid substrate is plastic.

14. A method of producing the array of claim 1 comprising: attaching the plurality of sequence specific polynucleotides at positions which are positionally distinguishable on the solid substrate.

15. A method of detecting hybridization of a target nucleic acid to the array of claim 1 comprising: contacting the target

US 6,576,424 B2

83

nucleic acid with the polynucleotides of the array under conditions permitting hybridization and detecting hybridization.

16. A method of comparing a target nucleic acid with a reference nucleic acid comprising:

- (a) contacting the target nucleic acid and the reference nucleic acid with the array of claim 1 under hybridization conditions,
- (b) determining a first pattern of hybridization with the target nucleic acid and a second pattern of hybridization with the reference nucleic acid, and
- (c) comparing the first and second patterns of hybridization to determine similarities or differences between the target nucleic acid and the reference nucleic acid.

17. A method for detecting nucleotide sequences in two or more collections of nucleic acids comprising:

- (a) providing the array of claim 1 including said positionally distinguishable sequence specific polynucleotides;
- (b) contacting said array of positionally distinguishable sequence specific polynucleotides with
 - (i) a first collection of labeled nucleic acids comprised of a nucleotide sequence that specifically binds to at least one known position of the array, and
 - (ii) at least a second collection of labeled nucleic acids comprised of a nucleotide sequence that specifically binds to at least one distinguishable position of the array; wherein the labels of the nucleic acids of the first and second collections are distinguishable from each other; and

- (c) detecting hybridization of the nucleic acids of the first and second collections to the polynucleotides of the array.

18. The method of claim 17, wherein the labels of the first and second nucleic acid collections are distinguishable fluorescent labels.

19. A method of identifying a sample containing nucleic acids comprising:

- (a) contacting the sample with the array of claim 1 under hybridization conditions to obtain a hybridization pattern, and
- (b) comparing the hybridization pattern with a reference database of hybridization patterns from the same array to identify the sample.

20. An array comprising a solid substrate and a plurality of positionally distinguishable polynucleotides attached to the solid substrate at a density of at least 100 polynucleotides per cm^2 ; wherein each of the attached polynucleotides comprises at least 50 nucleotides and has a predetermined nucleotide sequence.

21. The array of claim 20, wherein the positionally distinguishable polynucleotides are attached to the solid substrate within a square of less than 500 microns in dimension.

22. The array of claim 20, wherein each of the positionally distinguishable polynucleotides is attached to the solid substrate within a square of less than 100 microns in dimension.

23. The array of claim 20, wherein each of the positionally distinguishable polynucleotides is attached to the solid substrate within a square of less than 50 microns in dimension.

24. The array of claim 20, wherein each of the positionally distinguishable polynucleotides is attached to the solid substrate within a square of less than 10 microns in dimension.

25. The array of claim 20, wherein the density of positionally distinguishable polynucleotides attached to the solid substrate is at least 300 polynucleotides per cm^2 .

26. The array of claim 20, wherein the density of positionally distinguishable polynucleotides attached to the solid substrate is at least 1000 polynucleotides per cm^2 .

84

27. The array of claim 20, wherein the density of positionally distinguishable polynucleotides attached to the solid substrate is at least 3000 polynucleotides per cm^2 .

28. The array of claim 20, wherein the density of positionally distinguishable polynucleotides attached to the solid substrate is at least 10,000 polynucleotides per cm^2 .

29. The array of claim 20, wherein the density of positionally distinguishable polynucleotides attached to the solid substrate is at least 30,000 polynucleotides per cm^2 .

30. The array of claim 20, wherein the density of positionally distinguishable polynucleotides attached to the solid substrate is at least 100,000 polynucleotides per cm^2 .

31. The array of claim 20, wherein the plurality of attached polynucleotides comprises at least 3000 different polynucleotides.

32. The array of claim 20, wherein the solid substrate has a surface area of less than 4 square centimeters.

33. The array of claim 20, wherein the solid substrate has a surface area of less than 11 cm^2 .

34. The array of claim 20, wherein the solid substrate is glass.

35. The array of claim 20, wherein the solid substrate comprises a polymeric substrate.

36. The array of claim 20, wherein the solid substrate is plastic.

37. A method of producing the array of claim 20 comprising: attaching said polynucleotides at positions which are positionally distinguishable on the solid substrate.

38. A method of detecting hybridization of a target nucleotide acid to the array of claim 20 comprising: contacting the target nucleotide acid with the polynucleotides of the array under hybridization conditions and detecting hybridization.

39. A method of comparing a target nucleic acid with a reference nucleic acid comprising:

- (a) contacting the target nucleic acid and the reference nucleic acid with the array of claim 28 under conditions permitting hybridization,
- (b) determining a first pattern of hybridization with the target nucleic acid and a second pattern of hybridization with the reference nucleic acid, and
- (c) comparing the first and second patterns of hybridization to determine similarities or differences between the target nucleic acid and the reference nucleic acid.

40. A method for detecting nucleotide sequences in two or more collections of nucleic acids comprising:

- (a) providing the array of polynucleotides of claim 20;
- (b) contacting the array of polynucleotides with
 - (i) a first collection of labeled nucleic acids comprised of a nucleotide sequence that specifically binds to at least one known position of the array, and
 - (ii) at least a second collection of labeled nucleic acids comprised of a nucleotide sequence that specifically binds to at least one known position of the array;

wherein the labels of the nucleic acids of the first and at least second collections are distinguishable from each other; and

- (c) detecting hybridization of the nucleic acids of the first and at least second collections to the polynucleotides of the array thereby detecting nucleotide sequences in two or more collections of nucleic acids.

41. The method of claim 40, wherein the labels of the first and at least second nucleic acid collections are distinguishable fluorescent labels.

42. A method of identifying a sample containing nucleic acids comprising:

US 6,576,424 B2

85

- (a) contacting the sample with the array of claim 20 under hybridization conditions to obtain a hybridization pattern, and
- (b) comparing the hybridization pattern with a reference database of hybridization patterns from the same array to identify the sample.

43. A method for detecting nucleic acid sequences in two or more collections of nucleic acid molecules, the method comprising:

- (a) providing an array of nucleic acids bound to a solid surface, each of said nucleic acids comprising a determinable nucleic acid sequence;
- (b) contacting the array of nucleic acids with:
 - (i) a first collection of labeled nucleic acids, and
 - (ii) at least a second collection of labeled nucleic acids; wherein the labels on the first collection of nucleic acids and at least the second collection of nucleic acids are distinguishable from each other; and
- (c) detecting hybridization of the first collection of labeled nucleic acids and at least the second collection of labeled nucleic acids to nucleic acids on said array to thereby detect nucleic acid sequences in two or more collections of nucleic acid molecules.

44. The method of claim 43, wherein the array comprises an array of beads.

45. The method of claim 43, wherein the first and at least second labels are fluorescent labels.

46. A method of detecting differential expression of a plurality of nucleic acids in a first cell or population of cells with respect to expression of nucleic acids in a second cell or population of cells, said method comprising:

- adding nucleic acids from the two cells or populations of cells to an array of nucleic acids, under conditions that result in hybridization to nucleic acids on the array, the nucleic acids from the first and the second cell or cell populations having fluorescent labels that are distinguishable from one another; and

- examining the array by fluorescence under fluorescence excitation conditions to detect differential expression of the nucleic acids from the two cells or cell populations.

47. The method of claim 46, wherein the array of nucleic acids has at least 10^2 distinct nucleic acids on a surface area of about 1 cm^2 , each distinct nucleic acid being disposed at a separate, defined position in said array.

48. The method of claim 46, wherein the array of nucleic acids has at least 10^3 distinct nucleic acids on a surface area of about 1 cm^2 , each distinct nucleic acid being positionally distinguishable.

49. The method of claim 47 wherein the nucleic acids are individually attached to the array at specific positions.

50. The method of claim 43 wherein the nucleic acids are individually attached to the array at specific positions.

51. The method of claim 43, wherein the array of nucleic acids has at least 10^2 distinct nucleic acids on a surface area of about 1 cm^2 , each distinct nucleic acid being disposed at a separate, defined position in said array.

52. The method of claim 46, wherein the array of nucleic acids has at least 10^3 distinct nucleic acids on a surface area of about 1 cm^2 , each distinct nucleic acid being disposed at a separate, defined position in said array.

53. The method of claim 43 wherein the array comprises a support and a surface, the surface of the array comprising a different material than the support.

86

54. The method of claim 46, wherein the array comprises a support and a surface, the surface of the array comprising a different material than the support.

55. The method of claim 43 wherein the labels are selected from the group consisting of green and red fluorescent labels.

56. The method of claim 46, wherein the fluorescent labels are selected from the group consisting of green and red fluorescent labels.

57. A method for detecting nucleic acids in two or more collections of nucleic acid molecules, the method comprising:

- (a) providing an array of polynucleotides bound to a solid surface, each said polynucleotide comprising a determinable nucleic acid;
- (b) contacting the array of polynucleotides with:
 - (i) a first collection of labeled nucleic acids, and
 - (ii) at least a second collection of labeled nucleic acids; wherein the first and second labels are distinguishable from each other; and
- (c) detecting hybridization of the first collection of labeled nucleic acids and at least the second collection of labeled nucleic acids to polynucleotides of said array to thereby detect nucleic acids in two or more collections of nucleic acid molecules.

58. The method of claim 57, wherein the solid substrate comprises an array of beads.

59. The method of claim 57, wherein the first and second labels are fluorescent labels.

60. A method of detecting differential expression of each of a plurality of genes in a first cell with respect to expression of the same genes in a second cell, said method comprising:

- adding fluorescent labeled nucleic acids from the two cells to an array of polynucleotides complementary to a plurality of known genes of the two cells, under conditions that result in hybridization to polynucleotides on the array; and

- examining the array by fluorescence under fluorescence excitation conditions in which labeled nucleic acids from one of the cells that are hybridized to polynucleotides in the array give a distinct fluorescence emission color and labeled nucleic acid from the other cell which are hybridized to polynucleotides of said array give a different fluorescence emission color thereby enabling the detection of the differential expression of the genes in said first and second cells.

61. The method of claim 60, wherein the array of polynucleotides has at least 10^3 distinct polynucleotide in a surface area of about 1 cm^2 , each distinct polynucleotide being disposed at a separate, defined position in said array.

62. The method of claim 46, further comprising mixing the nucleic acids from the first and the second cell populations before they are added to the array.

63. The method of claim 46, further comprising simultaneously adding the nucleic acids from the first and the second cell populations to the array.

64. The method of claim 46, further comprising the sequential addition of the nucleic acids from the first and the second cell populations to the array.

* * * * *

EXHIBIT 5

(12) **United States Patent**
Dower et al.

(10) **Patent No.:** **US 7,056,666 B2**
(45) **Date of Patent:** **Jun. 6, 2006**

(54) **ANALYSIS OF SURFACE IMMOBILIZED
POLYMERS UTILIZING
MICROFLUORESCENCE DETECTION**

(75) Inventors: **William J. Dower**, Menlo Park, CA
(US); **Stephen P. A. Fodor**, Palo Alto,
CA (US)

(73) Assignee: **Affymetrix, Inc.**, Santa Clara, CA (US)

(*) Notice: Subject to any disclaimer, the term of this
patent is extended or adjusted under 35
U.S.C. 154(b) by 0 days.

(21) Appl. No.: **10/077,070**

(22) Filed: **Feb. 14, 2002**

(65) **Prior Publication Data**

US 2004/0029115 A9 Feb. 12, 2004

Related U.S. Application Data

(63) Continuation of application No. 08/829,893, filed on
Apr. 2, 1997, now abandoned, which is a continuation
of application No. 08/679,478, filed on Jul. 12, 1996,
now Pat. No. 5,902,723, which is a continuation of
application No. 07/626,730, filed on Dec. 6, 1990,
now Pat. No. 5,547,839.

(51) **Int. Cl.**
C12Q 1/68 (2006.01)
C12P 19/34 (2006.01)
C07H 21/02 (2006.01)

(52) **U.S. Cl.** **435/6**; 435/91.1; 435/91.2;
536/23.1; 536/24.3

(58) **Field of Classification Search** 435/6,
435/91.2; 536/23.1, 24.3
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

3,849,137 A 11/1974 Barznski et al.
3,862,056 A 1/1975 Hartman
4,072,576 A 2/1978 Arwin et al.
4,238,757 A 12/1980 Schenk
4,314,821 A 2/1982 Rice
4,339,528 A 7/1982 Goldman
4,405,771 A 9/1983 Jagur
4,444,878 A 4/1984 Paulus
4,444,892 A 4/1984 Malmros
4,517,338 A 5/1985 Urdea et al.
4,542,102 A 9/1985 Dattagupta et al.
4,555,490 A 11/1985 Merrill
4,562,157 A 12/1985 Lowe et al.
4,569,967 A 2/1986 Kornreich et al.
4,582,789 A 4/1986 Sheldon et al.
4,631,211 A 12/1986 Houghton
4,637,861 A 1/1987 Krull et al.
4,656,127 A 4/1987 Mundy
4,681,859 A 7/1987 Kramer
4,689,405 A 8/1987 Frank et al.

4,704,353 A 11/1987 Humphries et al.
4,713,326 A 12/1987 Dattagupta et al.
4,713,347 A 12/1987 Mitchell et al.
4,719,615 A 1/1988 Feyrer et al.
4,722,906 A 2/1988 Guire
4,762,881 A 8/1988 Kauer
4,777,019 A 10/1988 Dandekar
4,786,684 A 11/1988 Glass
4,794,150 A 12/1988 Steel
4,808,508 A 2/1989 Platzer
4,811,218 A * 3/1989 Hunkapiller et al. ... 364/413.01
4,822,566 A 4/1989 Newman
4,833,092 A 5/1989 Geysen
4,851,331 A 7/1989 Vary et al.
4,855,225 A 8/1989 Fung et al.
4,865,990 A 9/1989 Stead et al.
4,889,818 A 12/1989 Gelfand et al.
4,946,942 A 8/1990 Fuller et al.
4,962,037 A 10/1990 Jett et al.
4,965,188 A 10/1990 Mullis et al.
4,973,493 A 11/1990 Guire
4,979,959 A 12/1990 Guire
5,002,867 A 3/1991 Macevicz
5,026,840 A 6/1991 Dattagupta et al.
5,075,216 A 12/1991 Innis et al.
5,126,239 A 6/1992 Livak et al.
5,143,854 A 9/1992 Pirrung et al.
5,202,231 A 4/1993 Drmanac
5,492,806 A 2/1996 Drmanac et al.
5,503,980 A 4/1996 Cantor
5,525,464 A 6/1996 Drmanac et al.
5,527,681 A 6/1996 Holmes
5,547,839 A * 8/1996 Dower et al. 435/6
5,631,134 A 5/1997 Cantor
5,667,972 A 9/1997 Drmanac et al.
5,695,940 A 12/1997 Drmanac et al.

(Continued)

FOREIGN PATENT DOCUMENTS

DE 2242394 3/1974

(Continued)

OTHER PUBLICATIONS

"Affymax raises \$25 million to develop high-speed drug
discovery system," *Biotechnology News*, 10(3):7-8 (1990).

(Continued)

Primary Examiner—Jeffrey Fredman
(74) *Attorney, Agent, or Firm*—Townsend and Townsend
and Crew, LLP

(57) **ABSTRACT**

Means for simultaneous parallel sequence analysis of a large
number of biological polymer macromolecules. Apparatus
and methods may use fluorescent labels in repetitive chem-
istry to determine terminal manomers on solid phase immo-
bilized polymers. Reagents which specifically recognize
terminal manomers are used to label polymers at defined
positions on a solid substrate.

15 Claims, 10 Drawing Sheets

US 7,056,666 B2

Page 2

U.S. PATENT DOCUMENTS

5,753,439 A	5/1998	Smith et al.	
5,795,714 A	8/1998	Cantor et al.	
5,807,522 A	9/1998	Brown et al.	
5,849,878 A	12/1998	Cantor et al.	
5,902,723 A	5/1999	Dower et al.	
5,972,619 A	10/1999	Drmanac	
6,007,987 A	12/1999	Cantor et al.	
6,013,431 A *	1/2000	Soderlund et al.	435/5
6,018,041 A	1/2000	Drmanac	
6,025,136 A	2/2000	Drmanac	
6,054,270 A	4/2000	Southern	

FOREIGN PATENT DOCUMENTS

DE	3440141	5/1986
EP	088 636	9/1983
EP	127 438	12/1984
EP	245 662	11/1987
EP	288 310	10/1988
EP	319 012	6/1989
EP	328 256	8/1989
EP	373 203 B1 *	11/1989
EP	392 546	10/1990
EP	416 817 A2	3/1991
EP	607 151 B1	11/2002
GB	2233654 A	1/1991
JP	60-248699 A	12/1985
JP	63-084499 A	4/1988
WO	WO 84/03564	9/1984
WO	WO 86/00991	2/1986
WO	WO 86/06487	11/1986
WO	WO 89/10414 A1	11/1989
WO	WO 89/10977	11/1989
WO	WO 89/11548	11/1989
WO	WO 89/12819	12/1989
WO	WO 90/00887	2/1990
WO	WO 90/03382	4/1990
WO	WO 90/04652	5/1990
WO	WO 90/09455 A1	8/1990
WO	WO 90/11372 A1	10/1990
WO	WO 90/13666	11/1990
WO	WO 90/15070	12/1990
WO	WO 91/04266	4/1991
WO	WO 91/06678	5/1991
WO	WO 91/07087	5/1991
WO	WO 91/13075 A2	9/1991
WO	WO 93/17126	9/1993
WO	WO 95/09248	4/1995
WO	WO 98/31836	7/1998

OTHER PUBLICATIONS

Ajayaghosh et al., "Solid-Phase Synthesis of *N*-Methyl- and *N*-Ethylamides of Peptides Using Photolytically Detachable ((3-Nitro-4((alkylamino)methyl)benzamido)methyl)polystyrene Resin," *J. Org. Chem.*, 55(9):2826-2829 (1990).

Ajayaghosh et al., "Solid-phase synthesis of C-terminal peptide amides using a photoremovable α -methylphenacylamido anchoring linkage," *Proc. Ind. Natl. Sci. (Chem. Sci.)*, 100(5):389-396 (1988).

Ajayaghosh et al., "Polymer-supported Solid-phase Synthesis of C-Terminal Peptide *N*-Methylamides Using a Modified Photoremovable 3-Nitro-4-*N*-methylaminomethylpolystyrene Support," *Ind. J. Chem.*, 27B:1004-1008 (1988).

Ajayaghosh et al., "Polymer-Supported Synthesis of Protected Peptide Segments on a Photosensitive *o*-Nitro(α -Methyl)Bromobenzyl Resin," *Tetrahedron*, 44(21):6661-6666 (1988).

Amit et al., "Photosensitive Protecting Groups of Amino Sugars and Their Use in Glycoside Synthesis. 2-Nitrobenzyloxycarbonylamino and 6-Nitroveratryloxycarbonylamino Derivatives," *J. Org. Chem.*, 39(2):192-196 (1974).

Applied Biosystems, Model 431A Peptide Synthesizer User's manual, Sections 2 and 6, (Aug. 15, 1989).

Atherton et al., *Solid Phase Peptide Synthesis: A Practical Approach*, IRL Press, (1989), tbl. of cont., pp. vii-ix.

Bains et al., "A Novel Method for Nucleic Acid Sequence Determination," *J. Theor. Biol.*, 135:303-307 (1988).

Bains, W., "Alternative Routes Through the Genome," *Biotechnology*, 8:1251-1256 (1988).

Balachander et al., "Functionalized Siloxy-Anchored Monolayers with Exposed Amino, Azido, Bromo, or Cyano Groups," *Tetrahed. Ltrs.*, 29(44):5593-5594 (1988).

Barinaga, M., "Will 'DNA Chip' Speed Genome Initiative," *Science*, 253:1489 (1985).

Baum, R., "Fledgling firm targets drug discovery process," *Chem. Eng. News*, p. 10-11 (1990).

Biorad Chromatography Electrophoresis Immunochemistry Molecular Biology HPLC catalog M 1987 pp. 182.

Carrano et al., "A High-Resolution, Fluorescence-Based, Semiautomated Method for DNA Fingerprinting," *Genomics*, 4:129-136 (1989).

Chatterjee et al., "Inducible Alkylation of DNA Using an Oligonucleotide-Quinone Conjugate," *Am. J. Chem. Soc.*, 112:6397-6399 (1990).

Chetverin et al., "Oligonucleotide Arrays: New Concepts and Possibilities," *Biotechnology*, 12:1093-1099 (1994).

Chidgeavadze et al., "2', 3'-Dieoxy-3' aminonucleoside 5'-triphosphates are the terminators of DNA synthesis catalyzed by DNA polymerases," *Nuc. Acids Res.*, 12(3):1671-1686 (1984).

Chidgeavadze et al., "3'-Fluoro-2', 3'-dideoxyribonucleoside 5'-triphosphates: Terminators of DNA synthesis," *FEBS Letters*, 183(2):275-278 (1985).

Chien et al., "Deoxyribonucleic Acid Polymerase from the Extreme Thermophile *Thermus aquaticus*," *J. Bacteriol.*, 127:1550-1557 (1976).

Cimino et al., "Psoralens as Photoactive Probes of Nucleic Acid Structure and Function: Organic Chemistry, Photochemistry and Biochemistry," *Ann. Rev. Biochem.*, 54:1151-1193 (1985).

Corbett et al., "Reaction of Nitroso Aromatics with Glyoxylic Acid. A New Path to Hydroxamic Acids," *J. Org. Chem.*, 45:2834-2839 (1980).

Coulson et al., "Toward a physical map of the genome of the nematode *Caenorhabditis elegans*," *PNAS*, 83:7821-7825 (1986).

Craig et al., "Ordering of cosmid clones covering the Herpes simplex virus type I (HSV-1) genome: a test case for fingerprinting by hybridization," *Nuc. Acid. Res.*, 18(9):2653-2660 (1990).

Dower et al., "The Search for Molecular Diversity (II): Recombinant and Synthetic Randomized Peptide Libraries," *Ann. Rep. Med. Chem.*, 26:271-280 (1991).

Drmanac et al., "Subfragments as an informative characteristic of the DNA molecule—computer simulation," thesis submitted to University of Belgrade College of Natural Sciences and Mathematics, (1988).

Drmanac et al., "Partial Sequencing by Oligo-Hybridization Concept and Applications in Genome Analysis," *1st Int. Conf. Electrophor., Supercomp., Human Genome*, pp. 60-74 (1990).

US 7,056,666 B2

Page 3

- Drmanac et al., "Sequencing by Oligonucleotide Hybridization: A Promising Framework in Decoding of the Genome Program?," *1st Intl. Conf. Electrophor., Supercomp., Human Genome*, pp. 47-59 (1990).
- Drmanac et al., "Laboratory Methods, Reliable Hybridization of Oligonucleotides as Short as Six Nucleotides," *DNA and Cell Biol.*, 9(7):527-534 (1990).
- Drmanac et al., "Sequencing of Megabase Plus DNA by Hybridization: theory of the Method," *Genomics*, 4:114-128 (1989).
- Dulcey et al., "Deep UV Photochemistry of Chemisorbed Monolayers: Patterned Coplanar Molecular Assemblies," *Science*, 252:551-554 (1991).
- Duncan et al., "Affinity Chromatography of a Sequence-Specific DNA Binding Protein Using Teflon-Linked Oligonucleotides," *Analytical Biochemistry*, 169:104-108 (1988).
- Evans et al., "Physical mapping of complex genomes by cosmid multiplex analysis," *PNAS*, 86:5030-5034 (1989).
- Flanders et al., "A new interferometric alignment technique," *App. Phys. Ltrs.*, 31(7):426-429 (1977).
- Fodor et al., "Light-directed, Spatially Addressable Parallel Chemical Synthesis," *Science*, 251:767-773 (1991).
- Frank et al., "Simultaneous Multiple Peptide Synthesis Under Continuous flow Conditions on Cellulose Paper Discs as Segmental Solid Supports," *Tetrahedron*, 44(19):6031-6040 (1988).
- Frank et al., "Automation of DNA Sequencing Reactions and Related Techniques: A Workstation for Micromanipulation of Liquids," *Bio/Technology*, 6:1211-1212 (1988).
- Fuller et al., "Urethane-Protected Amino Acid N-Carboxy Anhydrides and Their Use in Peptide Synthesis," *J. Amer. Chem. Soc.*, 112(20):7414-7416 (1990).
- Furka et al., "General method for rapid synthesis of multicomponent peptide mixtures," *Int. J. Peptide Protein Res.*, 37:487-493 (1991).
- Furka et al., "Comucopia of Peptides by Synthesis," 14th int. Congress of Biochem. abst.# FR:013, Jul. 10-15, 1988 Prague, Czechoslovakia.
- Furka et al., "More Peptides by Less Labour," abst. 288, Int. Symp. Med. Chem., Budapest Hungary Aug. 15-19, 1988.
- Gait, eds., pp. 1-115 from *Oligonucleotide Synthesis: A Practical Approach*, IRL Press, (1984).
- Gazard et al., "Lithographic Technique Using Radiation-Induced Grafting of Acrylic Acid into Poly(Methyl Methacrylate) Films," *Polymer Engineering and Science*, 20(16):1069-1072 (1980).
- Gerard et al., "Influence on stability in *Escherichia coli* of the carboxy-terminal structure of cloned Moloney murine leukemia virus reverse transcriptase," *DNA*, 5(4):271-279 (1986).
- Getzoff et al., "Mechanisms of Antibody Binding to a Protein," *Science*, 235:1191-1196 (1987).
- Geysen et al., "Strategies for epitope analysis using peptide synthesis," *J. Immunol. Meth.*, 102:259-274 (1987).
- Geysen et al., "Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid," *PNAS*, 81:3998-4002 (1984).
- Geysen et al., "A synthetic strategy for epitope mapping," from *Peptides: Chem & Biol.*, Proc. of 10th Am. Peptide Symp., May 23-28, 1987, pp. 519-523, (1987).
- Geysen, "Antigen-antibody interactions at the molecular level: adventures in peptide synthesis," *Immunol. Today*, 6(12):364-369 (1985).
- Geysen et al., "Chemistry of Antibody Binding to a Protein," *Science*, 235:1184-1190 (1987).
- Geysen et al., "The delineation of peptides able to mimic assembled epitopes," 1986 CIBA Symp., pp. 130-149.
- Geysen et al., "Cognitive Features of Continuous Antigenic Determinants," *Mol. Recognit.*, 1(1):1-10 (1988).
- Geysen et al., "A Prio Ri Delineation of a Peptide Which Mimics A Discontinuous Antigenic Determinant," *Mol. Immunol.*, 23(7):709-715 (1986).
- Geysen et al., *Synthetic Peptides: Approaches to Biological Probes*, Alan R. Liss, Inc., pp. 19-20 (1989).
- Haridasan et al., "Peptide Synthesis using Photolytically Cleavable 2-Nitrobenzyloxycarbonyl Protecting Group," *Proc. Indian Natn. Sci. Acad.*, 53A(6):717-728 (1987).
- Ho et al., "Highly Stable Biosensor Using an Artificial Enzyme," *Anal. Chem.*, 59:536-537 (1987).
- Houts et al., "Reverse Transcriptase from Avian Myeloblastosis Virus," *J. Virol.*, 29(2):517-522 (1979).
- Ikehara et al., "The synthesis of Polynucleotides," *Advances in Carbohydrate Chem. & Biochem.*, 36:135-213 (1979).
- Innis et al., "DNA sequencing with *Thermus aquaticus* DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA," *PNAS*, 85:9436-9440 (1988).
- Jacobsen et al., "The N-terminal amino-acid sequences of DNA polymerase I from *Escherichia coli* and of the large and the small fragments obtained by a limited proteolysis," *Eur. J. Biochem.*, 45(2):623-627 (1974).
- Kaiser et al., "Peptide and Protein Synthesis by Segment Synthesis-Condensation," *Science*, 243:187-192 (1989).
- Kambra et al., "Optimization of Parameters in a DNA Sequenator Using Fluorescence Detection," *Bio/Tech*, 6:816-821 (1988).
- Karube, "Micro-biosensors based on silicon fabrication technology," chapter 25 from *Biosensors: Fundamentals and Applications*, Turner et al., eds., Oxford Publ., 1987, pp. 471-480 (1987).
- Khrapko et al., "An Oligonucleotide hybridization approach to DNA sequencing," *FEBS Lett.*, 256(1,2):118-122 (1989).
- Klenow et al., "Selective Elimination of the Exonuclease Activity of the Deoxyribonucleic Acid Polymerase from *Escherichia coli* B by Limited Proteolysis," *PNAS*, 65(2):168-175 (1970).
- Knight, P., "Materials and Methods/Microsequencers for Proteins and Oligosaccharides," *Bio/Tech.*, 7:1075-76 (1989).
- Kotewicz et al., "Cloning and overexpression of Moloney murine leukemia virus reverse transcriptase in *Escherichia coli*," *Gene*, 85:249-258 (1985).
- Kutateladze et al., "Analogues of nucleoside triphosphates with modified sugar residues as substrates for RNA polymerase," *Molekulyarnaya Biologiya*, 20(1):267-277 (1986).
- Lee et al., "synthesis of a Polymer Surface Containing Covalently Attached Triethoxysilane Functionality: Adhesion to Glass," *Macromolecules*, 21:3353-3356 (1988).
- Levy, M.F., "Preparing Additive Printed Circuits," *IBM Tech. Discl. Bull.*, 9(11):1473 (1967).
- Lieberman et al., "A Light source Smaller Than the Optical Wavelength," *Science*, 247:59-61 (1990).
- Little, P., "Clone maps made simple," *Nature*, 346:611-612 (1990).
- Lowe, C.R., "Biosensors," *Trends in Biotech.*, 2:59-65 (1984).
- Lowe, C.R., "An Introduction to the Concepts and Technology of Biosensors," *Biosensors*, 1:3-16 (1985).

US 7,056,666 B2

Page 4

- Lowe, C. R., *Biotechnology and Crop Improvement and Protection*, BCPC Publications, pp. 131-138 (1986).
- Lowe et al., "Solid-Phase Optoelectronic Biosensors," *Methods in Enzymology*, 137:338-347 (1988).
- Lowe, C.R., "Biosensors," *Phil. Tran. R. Soc. Lond.*, 324:487-496 (1989).
- Lysov et al., "A new method for determining the DNA nucleotide sequence by hybridization with oligonucleotides," *Doklady Biochem.*, 303(1-6):436-438 (1989).
- Maxam et al., "Sequencing End-Labeled DNA with Base-Specific Chemical Cleavages," *Meth. Enzymol.*, 65:499-560 (1980).
- McCray et al., "Properties and Uses of Photoreactive Caged Compounds," *Ann. Rev. Biophys. Biophys. Chem.*, 18:239-270 (1989).
- McGillis, VLSI Technology, Sze, eds., Chapter 7, "Lithography," pp. 267-301 (1983).
- Merrifield, R.B., "Solid Phase peptide Synthesis. I. The Synthesis of a Tetrapeptide," *J.Am.Chem.Soc.*, 85:2149-2154 (1963).
- Michiels et al., "Molecular approaches to genome analysis: a strategy for the construction of ordered overlapping clone libraries" *CABIOS*, 3(3):203-10 (1987).
- Morita et al., "Direct pattern fabrication on silicone resin by vapor phase electron beam polymerization," *J.Vac.Sci. Technol.*, B1(4):1171-1173 (1983).
- Nelson et al., "A new and versatile reagent for incorporating multiple primary aliphatic amines into synthetic oligonucleotides," *Nuc. Acids Res.*, 17(18):7179-7186 (1989).
- Nossal, N.G., "DNA synthesis on a Double-stranded DNA template by the T4 Bacteriophage DNA polymerase and the T4 Gene 32 DNA Unwinding Protein," *J.Biol. Chem.*, 249(17):5668-5676 (1974).
- Ohtsuka et al., "Studies on transfer ribonucleic acids and related compounds. IX Ribonucleic oligonucleotide synthesis using a photosensitive 0-nitrobenzyl protection at the 2'-hydroxyl group," *Nuc.Acids.Res.* 1(10):1351-1357 (1974).
- Olson et al., "Random-clone strategy for genomic restriction mapping in yeast," *PNAS*, 83:7826-7830 (1986).
- Parsons, B.J., "Psoralen Photochemistry," *Photochem. Photobiol.*, 32:813-821 (1980).
- Patchornik et al., "Photosensitive Protecting Groups," *J.Am. Chem.Soc.*, 92(21):6333-6335 (1970).
- Pevzner, P.A., "1-Tuple DNA Sequencing: Computer Analysis," *J. Biomol. Struct. Dynam.*, 7(1):63-69 (1989).
- Pfeifer et al., "Genomic Sequencing and Methylation Analysis by Ligation Mediated PCR," *Science*, 246:810-813 (1989).
- Pidgeon et al., "Immobilized Artificial Membrane Chromatography: Supports Composed of Membrane Lipids," *Anal. Biochem.*, 176:36-47 (89).
- Poustka et al., "Molecular Approaches to Mammalian Genetics," Cold Spring Harbor Symposia on Quantitative Biology, 51:131-139 (1986).
- Prober et al., "A System for Rapid DNA Sequencing with Fluorescent Chain-Terminating Dinucleotides," *Science*, 238:336-341 (1987).
- Purushothaman et al., "Synthesis of 4,5-diarylimidazole-2-thiones and their photoconversion to bis(4,5-diarylimidazol-2-yl) sulphides," *Ind. J. Chem.*, 29B:18-21 (1990).
- Roberts, L., "A Sequencing Reality Check," *Science*, 242:1245 (1988).
- Rodda et al., "The Antibody Response to Myoglobin-I. Systematic Synthesis of Myoglobin Peptides Reveals Location and Substructure of Species-Dependent Continuous Antigenic Determinants," *Mol. Immunol.*, 23(6):603-610 (1986).
- Ross et al., "Interstrand Crosslinks due to 4, 5', 8-trimethylpsoralen and Near Ultraviolet Light in Specific Sequences of Animal DNA," *J. Mol. Biol.*, 201:339-351 (1988).
- Ruth et al., "Nucleoside Analogues with Clinical Potential in Antivirus Chemotherapy," *Mol. Pharm.*, 20:415-422 (1981).
- Saiki et al., "Introducing AmpliTaq DNA polymerase," *Amplifications*, 1:4-6 (1989).
- Sanger et al., "A Rapid Method for Determining Sequences in DNA by Primed Synthesis with DNA Polymerase," *J. Mol. Biol.*, 94:441-448 (1975).
- Sanger et al., "DNA sequencing with chain-terminating inhibitors," *PNAS*, 74(12):5463-5467 (1977).
- Second College edition of the American Heritage Dictionary, p. 522.
- Seed, B., "Diazotizable arylamine cellulose papers for the coupling and hybridization of nucleic acids," *Nuc. Acids Res.*, 10(5):1799-1810 (1982).
- Smith et al., "The synthesis of oligonucleotides containing an aliphatic amino group at the 5' terminus: Synthesis of fluorescent DNA primers for use in DNA sequence analysis," *Nuc. Acids Res.*, 13(7):2399-2412 (1985).
- Smith et al., "Fluorescence detection in automated DNA sequence analysis," *Nature*, 321:674-679 (1986).
- Smith et al., "A Novel Method for Delineating Antigenic Determinants: Peptide Synthesis and Radioimmunoassay Using the Same Solid Support," *Immunochemistry*, 14:565-568 (1977).
- Song et al., "Photochemistry and photobiology of psoralens," *Photochem. Photobiol.*, 29:1177-1197 (1979).
- Steuber et al., "Synthesis and Photolytic Cleavage of Bovine Insulin B22-30 on a Nitrobenzoylglycyl-polyethylene glycol support," *Chemical abstracts*, 100(17):700, abstract No. 139591v (1984).
- Stuber et al., "Synthesis and photolytic cleavage of bovine insulin B₂₂₋₃₀ on a nitrobenzoylglycyl-poly (ethylene glycol) support," *Intl. J. Peptide Protein Res.*, 22:277-283 (1983).
- Tabor et al., "Selective Oxidation of the Exonuclease Domain of Bacteriophage T7 DNA Polymerase," *J. Biol. Chem.*, 262(32):15330-15333 (1987).
- Tabor et al., "DNA sequence analysis with a modified bacteriophage T7 DNA polymerase," *PNAS*, 84:4767-4771 (1987).
- Tsugita et al., "Sensitization of Edman Amino Acid Derivatives Using the Fluorescent Reagent 4-Aminofluorescein," *J. Biochem.*, 106:60-65 (1989).
- Turner et al., "Photochemical Activation of Acylated α -Thrombin," *J. Am. Chem. Soc.*, 109:1274-1275 (1987).
- Weising et al., "Foreign genes in plants: Transfer, structure, expression, and applications," *Annu. Rev. Genet.*, 22:421-477 (1988).
- Wiesehahn et al., "DNA unwinding induced by photoaddition of psoralen derivatives and determination of dark binding equilibrium constants by gel electrophoresis," *PNAS*, 75:2703-2707 (1978).
- Wittman-Liebold, eds., *Methods in Protein Sequence Analysis*, from Proceedings of the 7th Int'l Conf., Berlin, Germany, Jul. 3-8, 1988, table of contents, pp. xi-xx* (1989).

US 7,056,666 B2

Page 5

Wood et al., "Base composition-independent hybridization in tetramethylammonium chloride: A method for oligonucleotide screening of highly complex gene libraries," *PNAS*, 82:1585-1588 (1985).

Ye et al., "Heat-stable DNA polymerase I large fragment resolves hairpin structure in DNA sequencing," *Scientia Sinica (Series B)*, 30(5):503-506 (1987).

Yosomiya et al., "Performance, Glass fiber Having Isocyanate Group on the Surface. Preparation and Reaction with Amino Acid," *Polymer Bulletin*, 12:41-48 (1984).

Zehavi et al., "Light-Sensitive Glycosides. I. 6-Nitroveratryl β -D-Glucopyranoside and 2-Nitrobenzyl β -D-Glucopyranoside," *J. Org. Chem.*, 37(14):2281-2285 (1972).

* cited by examiner

U.S. Patent

Jun. 6, 2006

Sheet 1 of 10

US 7,056,666 B2

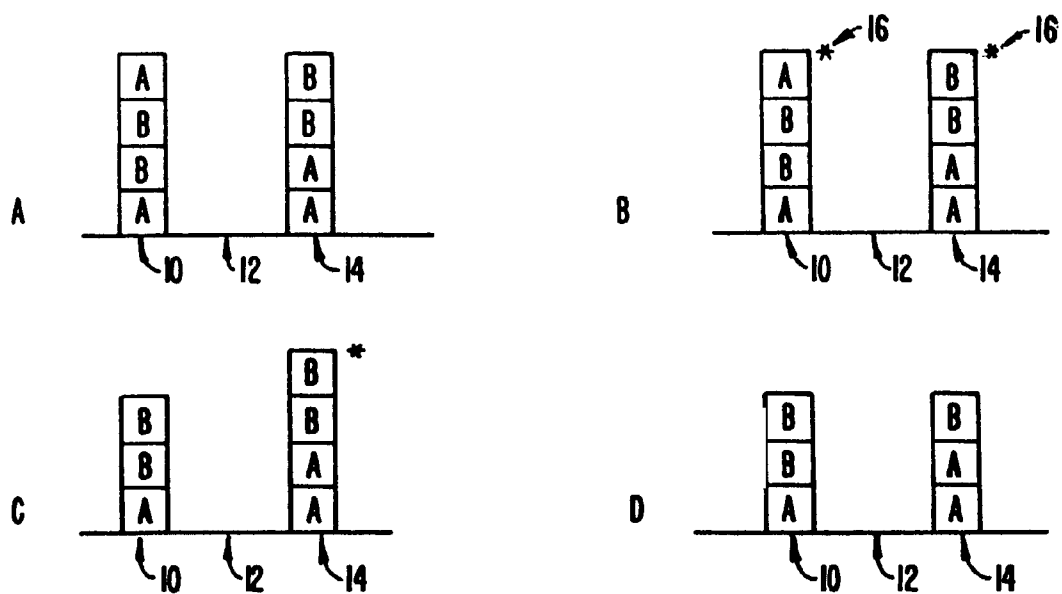


FIG. 1.

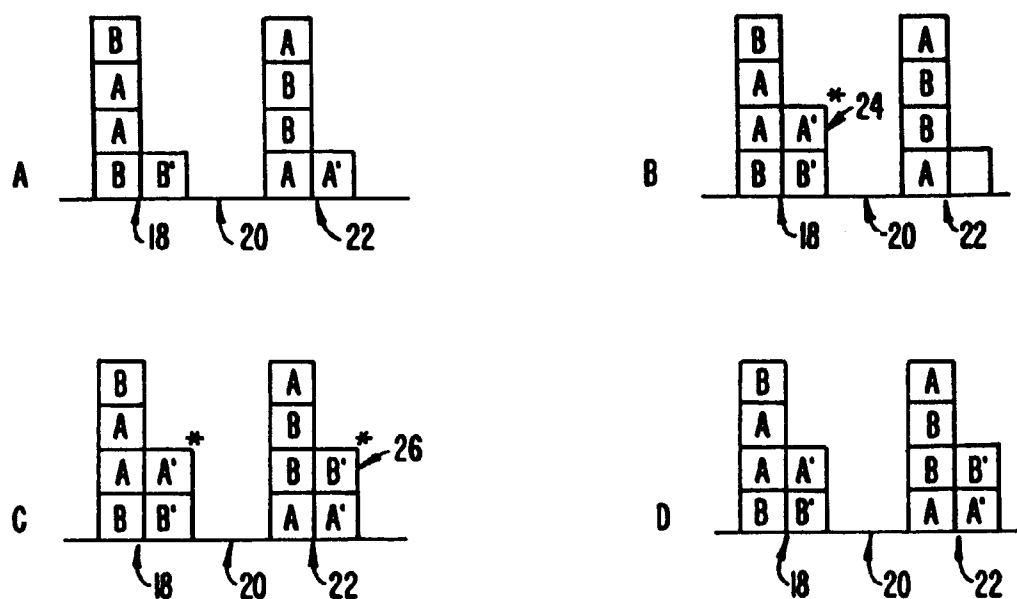


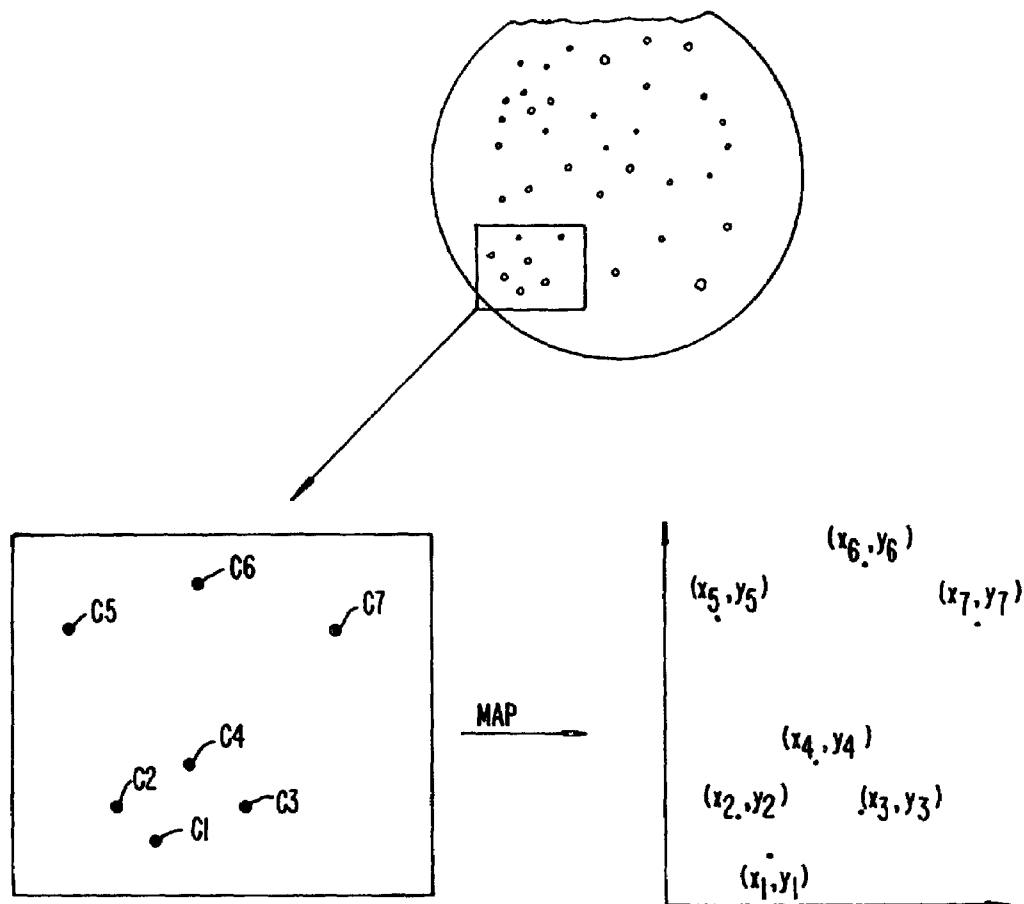
FIG. 2.

U.S. Patent

Jun. 6, 2006

Sheet 2 of 10

US 7,056,666 B2



CLUSTERS LOCALIZED AT POSITIONS

C_1	(x_1, y_1)
C_2	(x_2, y_2)
C_3	(x_3, y_3)
C_4	(x_4, y_4)
C_5	(x_5, y_5)
C_6	(x_6, y_6)
C_7	(x_7, y_7)
\vdots	\vdots
C_n	(x_n, y_n)

FIG. 3.

U.S. Patent

Jun. 6, 2006

Sheet 3 of 10

US 7,056,666 B2

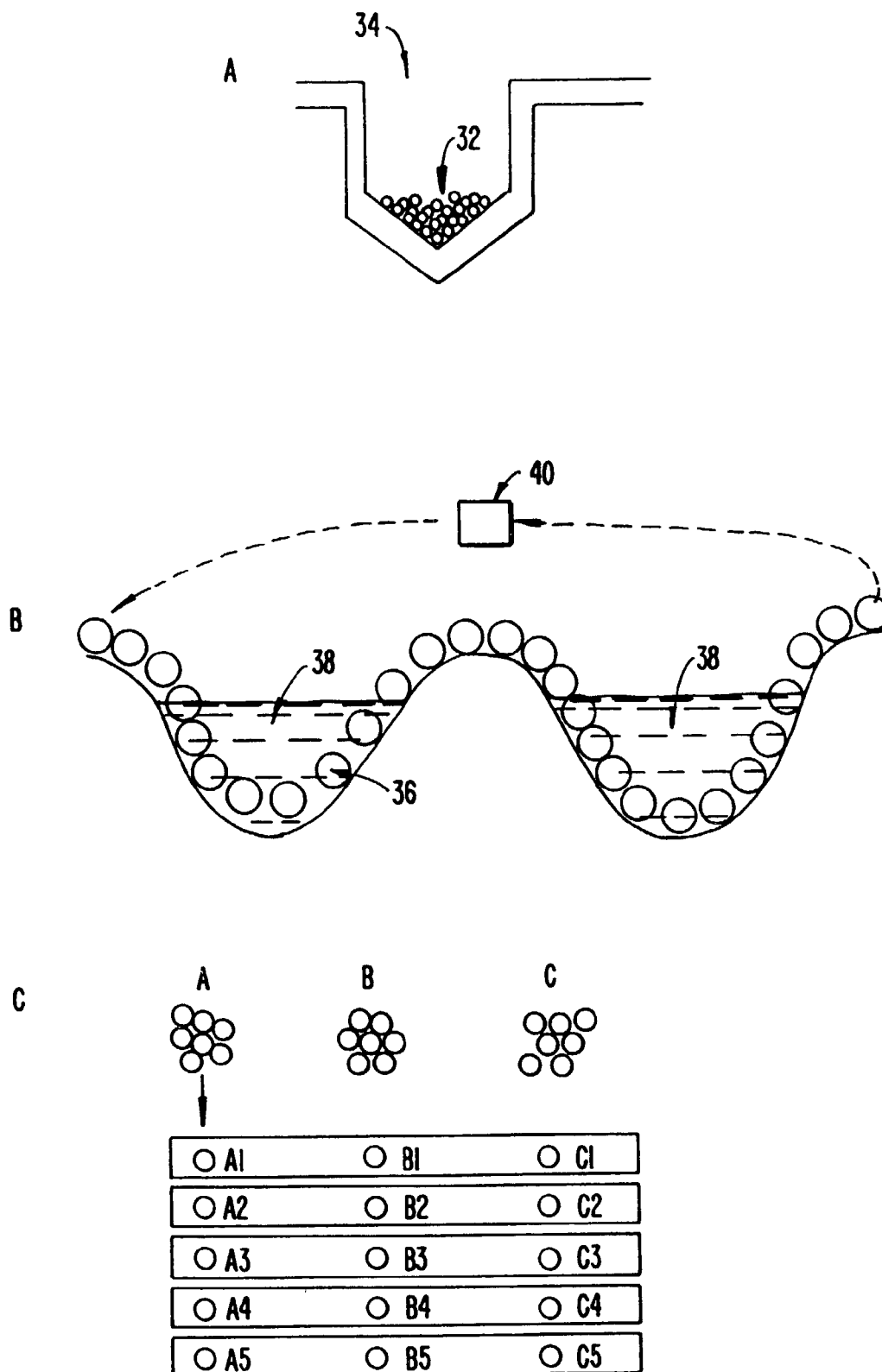


FIG. 4.

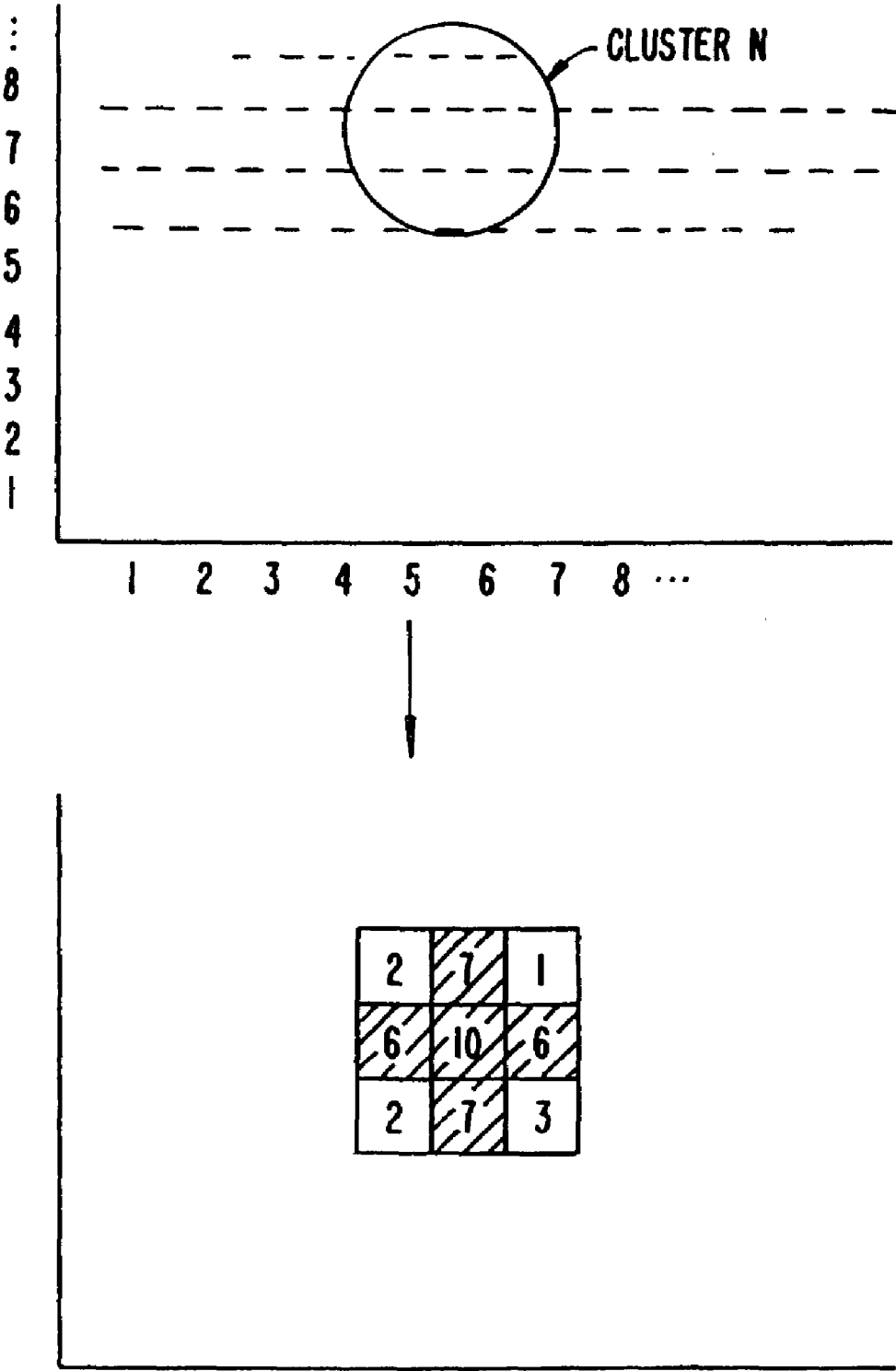


FIG. 5.

U.S. Patent

Jun. 6, 2006

Sheet 5 of 10

US 7,056,666 B2

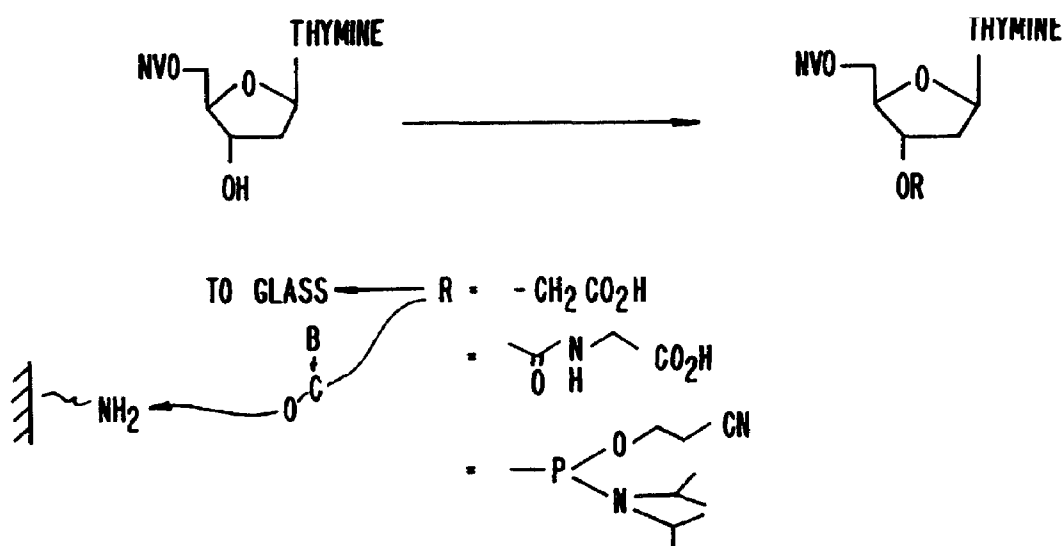


FIG. 6.

U.S. Patent

Jun. 6, 2006

Sheet 6 of 10

US 7,056,666 B2

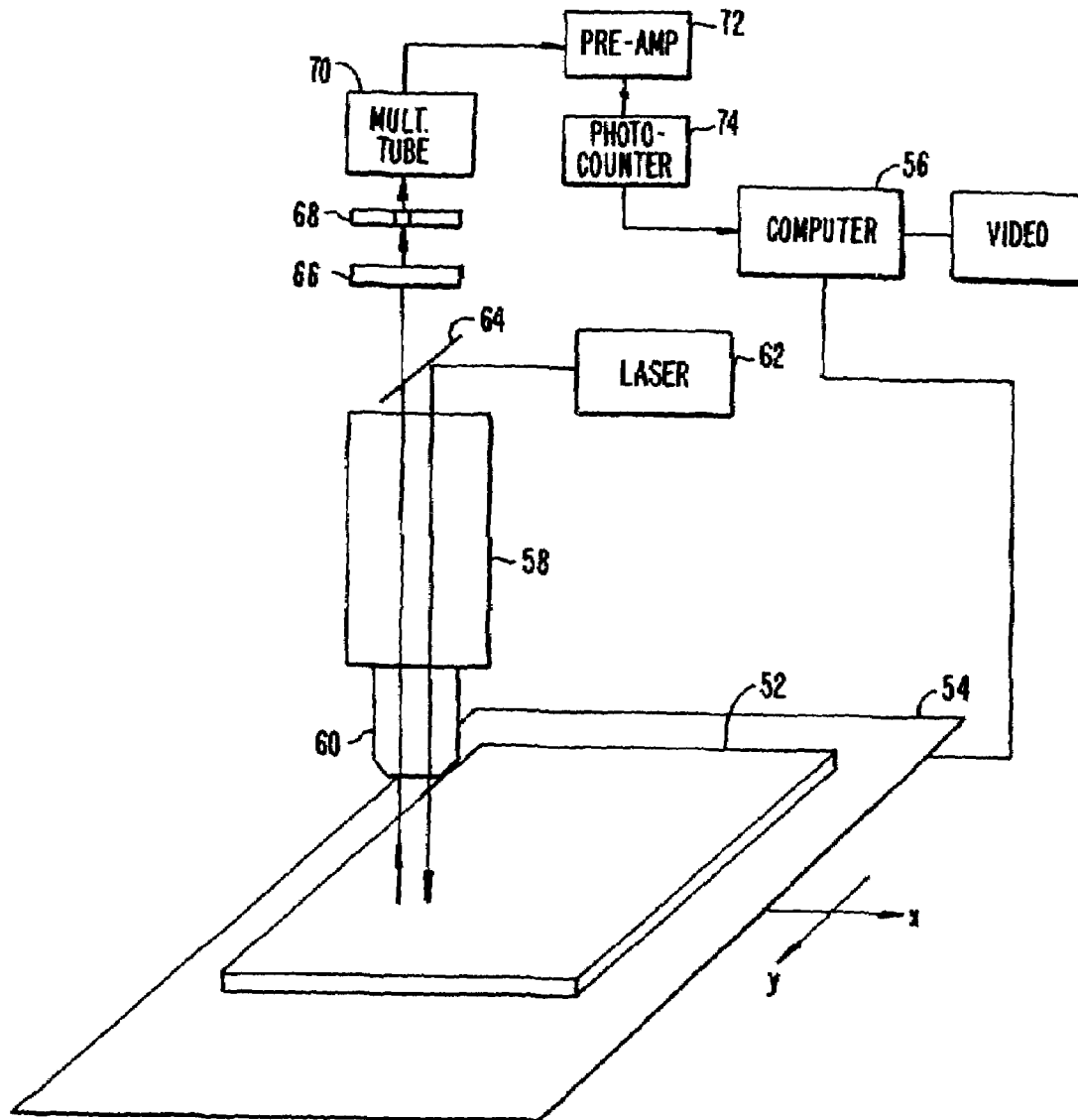


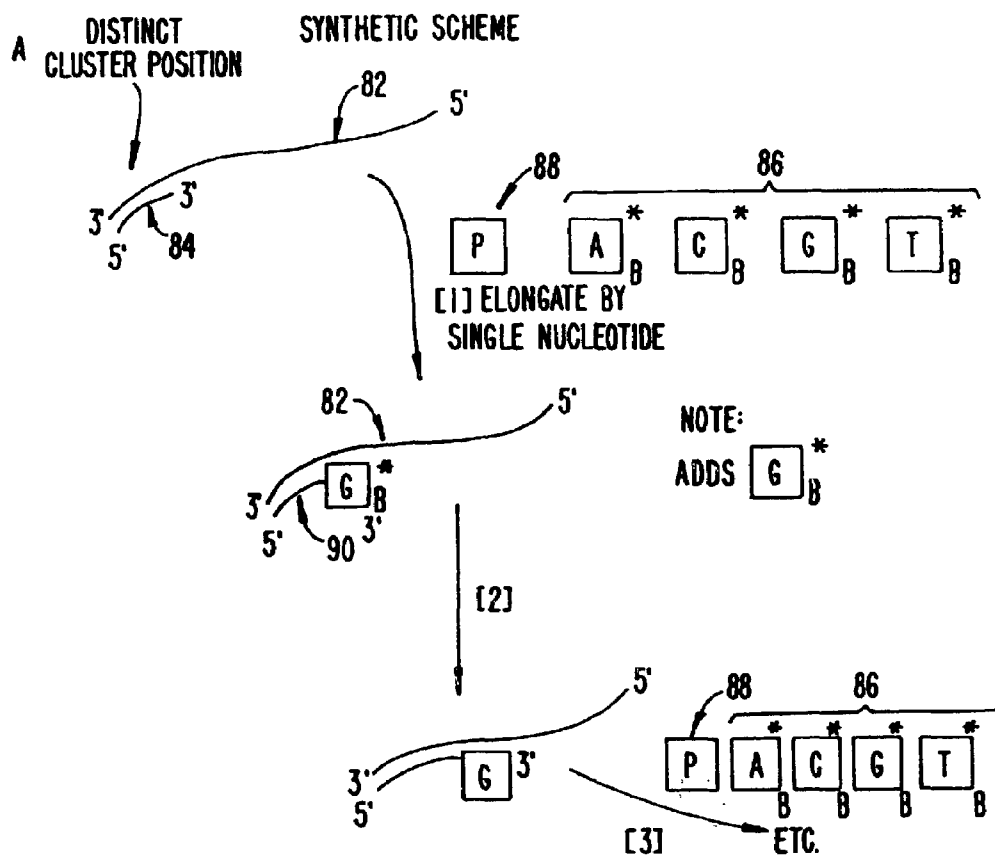
FIG. 7.

U.S. Patent

Jun. 6, 2006

Sheet 7 of 10

US 7,056,666 B2



B

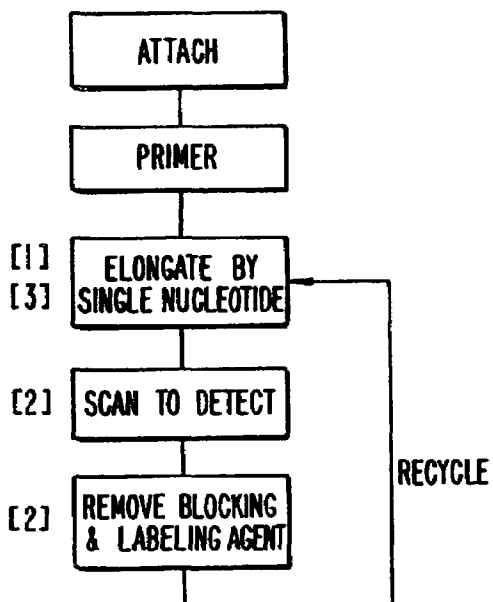


FIG. 8.

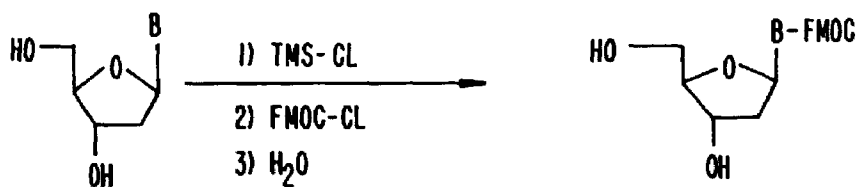
U.S. Patent

Jun. 6, 2006

Sheet 8 of 10

US 7,056,666 B2

PATHWAY TO PROTECTED NUCLEOTIDES



PREFERRED PATHWAY TO BASE PROTECTION AND FUNCTIONALIZATION

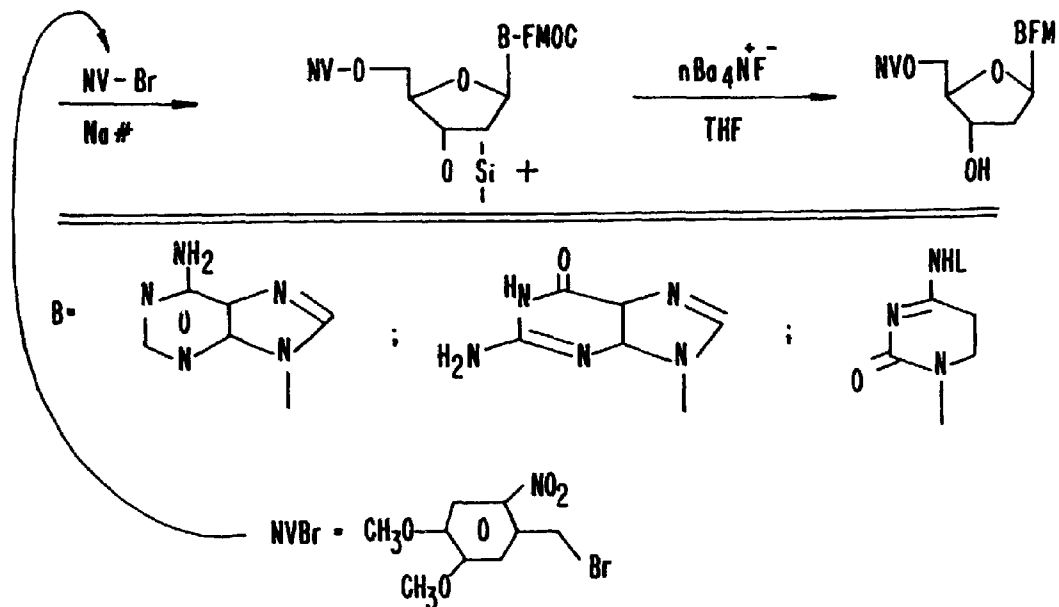
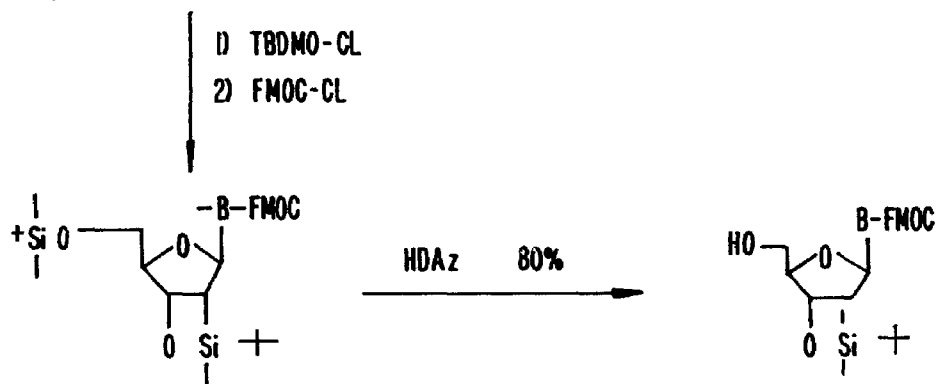


FIG. 9.

U.S. Patent

Jun. 6, 2006

Sheet 9 of 10

US 7,056,666 B2

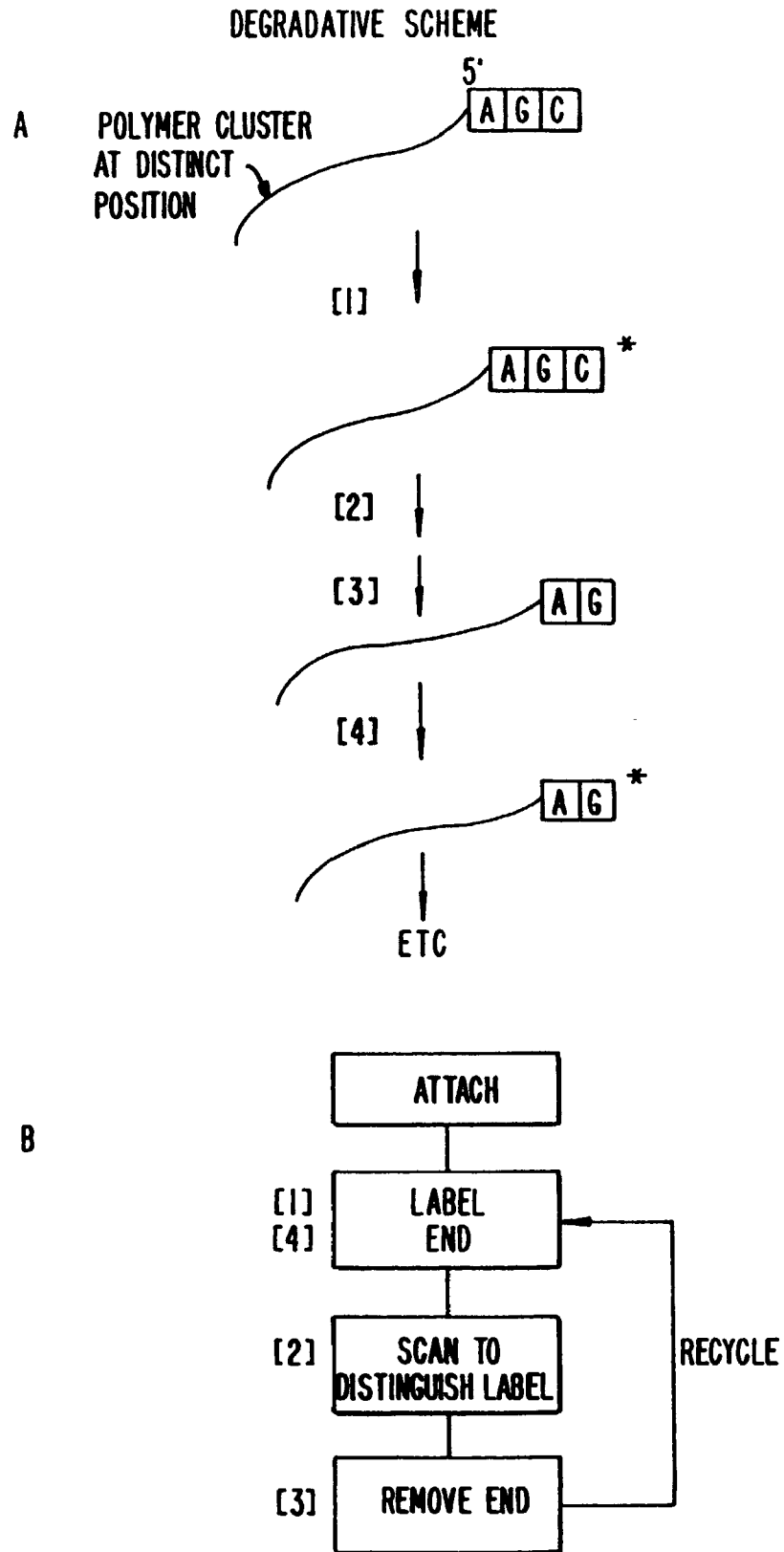


FIG. 10.

U.S. Patent

Jun. 6, 2006

Sheet 10 of 10

US 7,056,666 B2

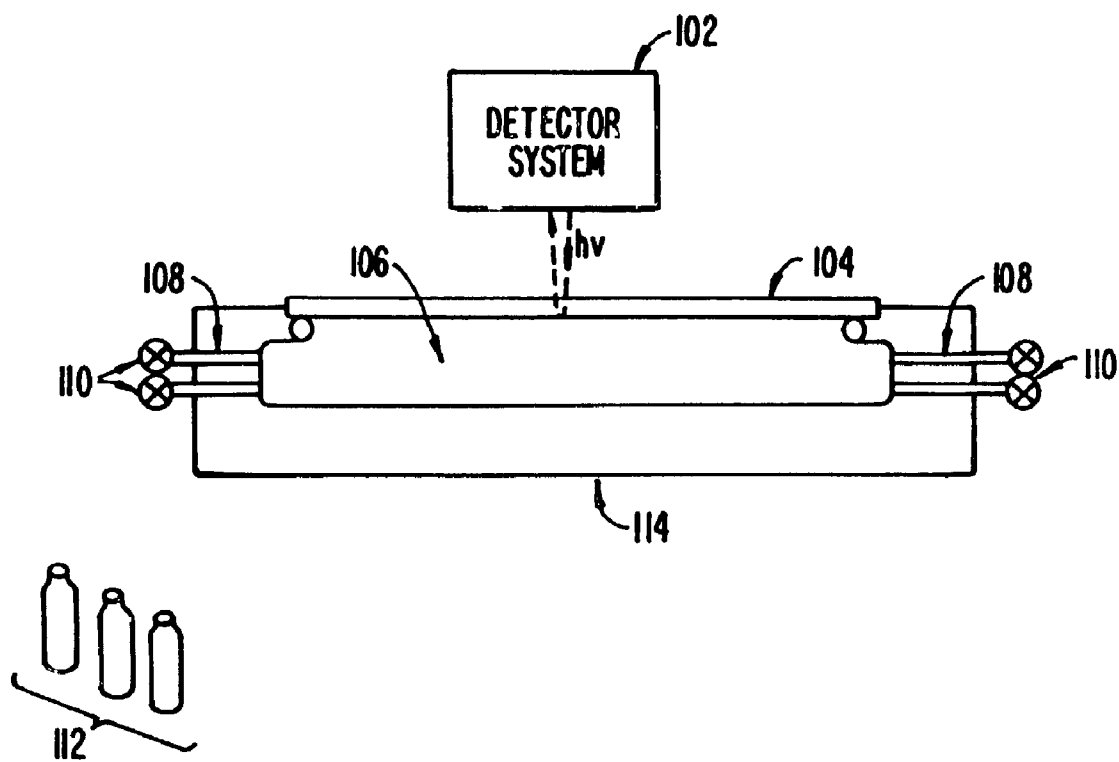


FIG. II.

US 7,056,666 B2

1

ANALYSIS OF SURFACE IMMOBILIZED POLYMERS UTILIZING MICROFLUORESCENCE DETECTION

The present application is a continuation of U.S. Ser. No. 08/829,893, filed Apr. 2, 1997 (now abandoned), which is a continuation of U.S. Ser. No. 08/679,478 (now U.S. Pat. No. 5,902,723), filed Jul. 12, 1996, which is a continuation of U.S. Ser. No. 07/626,730 (now U.S. Pat. No. 5,547,839), filed Dec. 6, 1990.

BACKGROUND OF THE INVENTION

The present invention relates to the determination of the sequences of polymers immobilized to a substrate. In particular, one embodiment of the invention provides a method and apparatus for sequencing many nucleic acid sequences immobilized at distinct locations on a matrix surface. The principles and apparatus of the present invention may be used, for example, also in the determination of sequences of peptides, polypeptides, oligonucleotides, nucleic acids, oligosaccharides, phospholipids and other biological polymers. It is especially useful for determining the sequences of nucleic acids and proteins.

The structure and function of biological molecules are closely interrelated. The structure of a biological polymer, typically a macromolecule, is generally determined by its monomer sequence. For this reason, biochemists historically have been interested in the sequence characterization of biological macromolecule polymers. With the advent of molecular biology, the relationship between a protein sequence and its corresponding encoding gene sequence is well understood. Thus, characterization of the sequence of a nucleic acid encoding a protein has become very important.

Partly for this reason, the development of technologies providing the capability for sequencing enormous amounts of DNA has received great interest. Technologies for this capability are necessary for, for example, the successful completion of the human genome sequencing project. Structural characterization of biopolymers is very important for further progress in many areas of molecular and cell biology.

While sequencing of macromolecules has become extremely important, many aspects of these technologies have not advanced significantly over the past decade. For example, in the protein sequencing technologies being applied today the Edman degradation methods are still being used. See, e.g., Knight (1989) "Microsequencers for Proteins and Oligosaccharides," *Bio/Technol.* 7:1075-1076. Although advanced instrumentation for protein sequencing has been developed, see, e.g., Frank et al. (1989) "Automation of DNA Sequencing Reactions and Related Techniques: A Work Station for Micromanipulation of Liquids," *Bio/Technol.* 6:1211-1213, this technology utilizes a homogeneous and isolated protein sample for determination of removed residues from that homogeneous sample.

Likewise, in nucleic acid sequencing technology, three major methods for sequencing have been developed, of which two are commonly used today. See, e.g., Sambrook et al. (1989) *Molecular Cloning: A Laboratory Manual* (2d Ed.) Vols. 1-3, Cold Spring Harbor Press, New York, which is hereby incorporated herein by reference. The first method was developed by Maxam and Gilbert. See, e.g., Maxam and Gilbert (1980) "Sequencing End-Labeled DNA with Base-Specific Chemical Cleavages," *Methods in Enzymol.* 65:499-560, which is hereby incorporated herein by reference. The polymer is chemically cleaved with a series of base-specific cleavage reagents thereby generating a series

2

of fragments of various lengths. The various fragments, each resulting from a cleavage at a specific base, are run in parallel on a slab gel which resolves nucleic acids which differ in length by single nucleotides. A protein specific label allows detection of cleavages at all nucleotides relative to the position of the label.

This separation requires high resolution electrophoresis or some other system for separating nucleic acids of very similar size. Thus, the target nucleic acid to be sequenced must usually be initially purified to near homogeneity.

Sanger and Coulson devised two alternative methods for nucleic acid sequencing. The first method, known as the plus and minus method, is described in Sanger and Coulson (1975) *J. Mol. Biol.* 94:441-448, and has been replaced by the second method. Subsequently, Sanger and Coulson developed another improved sequencing method known as the dideoxy chain termination method. See, e.g., Sanger et al. (1977) "DNA Sequencing with Chain-Termination Inhibitors," *Proc. Natl. Acad. Sci. USA* 74:5463-5467, which is hereby incorporated herein by reference. This method is based on the inability of 2', 3' dideoxy nucleotides to be elongated by a polymerase because of the absence of a 3' hydroxyl group on the sugar ring, thus resulting in chain termination. Each of the separate chain terminating nucleotides are incorporated by a DNA polymerase, and the resulting terminated fragment is known to end with the corresponding dideoxy nucleotide. However, both of the Sanger and Coulson sequencing techniques usually require isolation and purification of the nucleic acid to be sequenced and separation of nucleic acid molecules differing in length by single nucleotides.

Both the polypeptide sequencing technology and the oligonucleotide sequencing technologies described above suffer from the requirement to isolate and work with distinct homogeneous molecules in each determination.

In the polypeptide technology, the terminal amino acid is sequentially removed and analyzed. However, the analysis is dependent upon only one single amino acid being removed, thus requiring the polypeptide to be homogeneous.

In the case of nucleic acid sequencing, the present techniques typically utilize very high resolution polyacrylamide gel electrophoresis. This high resolution separation uses both highly toxic acrylamide for the separation of the resulting molecules and usually very high voltages in running the electrophoresis. Both the purification and isolation techniques are highly tedious, time consuming and expensive processes.

Thus, a need exists for the capability to simultaneously sequence many biological polymers without individual isolation and purification. Moreover, dispensing with the need to individually perform the high resolution separation of related molecules leads to greater safety, speed, and reliability. The present invention solves these and many other problems.

SUMMARY OF THE INVENTION

The present invention provides the means to sequence hundreds, thousands or even millions of biological macromolecules simultaneously and without individually isolating each macromolecule to be sequenced. It also dispenses with the requirement, in the case of nucleic acids, to separate the products of the sequencing reactions on dangerous polyacrylamide gels. Adaptable to automation, the cost and effort required in sequence analysis will be dramatically reduced.

This invention is most applicable, but not limited, to linear macromolecules. It also provides specific reagents for

US 7,056,666 B2

3

sequencing both oligonucleotides and polypeptides. It provides an apparatus for automating the processes described herein.

The present invention provides methods for determining the positions of polymers which terminate with a given monomer, where said polymers are attached to a surface having a plurality of positionally distinct polymers attached thereto, said method comprising the steps of:

labeling a terminal monomer in a monomer type specific manner; and

scanning said surface, thereby determining the positions of said label. In one embodiment, the polymers are polynucleotides, and usually the labeling of the terminal marker comprises incorporation of a labeled terminal monomer selected from the group of nucleotides consisting of adenine, cytidine, guanidine and thymidine.

An alternative embodiment provides methods for concurrently determining which subset of a plurality of positionally distinct polymers attached to a solid substrate at separable locations terminates with a given terminal subunit, said method comprising the steps of:

mixing said solid substrate with a solution comprising a reagent, which selectively marks positionally distinct polymers which terminate with said given terminal subunit; and

determining with a detector which separable locations are marked, thereby determining which subset of said positionally distinct polymers terminated with said given terminal subunit. In one version, the solution comprises a reagent which marks the positionally distinct polymer with a fluorescent label moiety. In another version the terminal subunit is selected from the group consisting of adenosine, cytosine, guanosine, and thymine.

Methods are also provided for determining which subset of a plurality of primer polynucleotides have a predetermined oligonucleotide, wherein the polynucleotides are complementary to distinctly positioned template strands which are attached to a solid substrate, said method comprising the steps of:

selectively marking said subset of primer polynucleotides having the predetermined oligonucleotide; and

detecting which polynucleotides are marked. In one embodiment, the oligonucleotide subunit is a single nucleotide; in another the marking comprises elongating said primer with a labeled nucleotide which is complementary to a template; and in a further embodiment the marking step uses a polymerase and a blocked and labeled adenine.

The invention embraces methods for concurrently obtaining sequence information on a plurality of polynucleotides by use of a single label detector, said method comprising the steps of:

attaching a plurality of positionally distinct polynucleotides to a solid substrate at separable locations;

labeling said plurality of polynucleotides with a terminal nucleotide specific reagent, said label being detectable using said label detector;

determining whether said specific labeling reagent has labeled each separable location. Often, the labeling is performed with reagents which can distinguishably label alternative possible nucleotide monomers. One embodiment uses four replica substrates each of which is labeled with a specific labeling reagent for adenine, cytosine, guanine, or thymine. Usually, the labeling and

4

determining steps are performed in succession using reagents specific for each of adenine, cytosine, guanine, and thymine monomers.

An alternative embodiment provides methods for concurrently obtaining sequence information on a plurality of polynucleotides, said method comprising the steps of:

attaching distinct polynucleotides to a plurality of distinct solid substrates;

labeling said plurality of solid substrates with a terminal nucleotide specific labeling reagent; and

determining whether said specific labeling reagent has labeled each distinct substrate. The method can be performed using in a continuous flow of distinct solid substrates through a reaction solution.

A method is provided for simultaneously sequencing a plurality of polymers made up of monomer units, said plurality of polymers attached to a substrate at definable positions, said method comprising the steps of:

mixing said substrate with a reagent which specifically recognizes a terminal monomer, thereby providing identification among various terminal monomer units; and

scanning said substrate to distinguish signals at definable positions on said substrate; and

correlating said signals at defined positions on said substrate to provide sequential series of sequence determinations. Often, the plurality of polymers are synthesized by a plurality of separate cell colonies, and the polymers may be attached to said substrate by a carbonyl linkage. In one embodiment, the polymers are polynucleotides, and often the substrate comprises silicon. The scanning will often identify a fluorescent label. In one embodiment, the reagent exhibits specificity of removal of terminal monomers, in another, the reagent exhibits specificity of labeling of terminal monomers.

The invention also embraces methods for sequencing a plurality of distinctly positioned polynucleotides attached to a solid substrate comprising the steps of:

hybridizing complementary primers to said plurality of polynucleotides;

elongating a complementary primer hybridized to a polynucleotide by adding a single nucleotide; and

identifying which of said complementary primers have incorporated said nucleotide. In some versions, the elongating step is performed simultaneously on said plurality of polynucleotides linked to said substrate. Typically, the substrate is a two dimensional surface and the identifying results from a positional determination of the complementary primers incorporating the single defined nucleotide. A silicon substrate is useful in this method.

Methods, are provided where the linking is by photocrosslinking polynucleotide to said complementary primer, where said primer is attached to said substrate. The elongating will be often catalyzed by a DNA dependent polymerase. In various embodiments, a nucleotide will have a removable blocking moiety to prevent further elongation, e.g., NVOC.

A nucleotide with both a blocking moiety and labeling moiety will be often used.

A further understanding of the nature and advantages of the invention herein may be realized by reference to the remaining portions of the specification and the attached drawings.

US 7,056,666 B2

5

BRIEF DESCRIPTION OF THE FIGURES

FIG. 1 illustrates a simplified and schematized embodiment of a degradative scheme for polymer sequencing.

FIG. 2 illustrates a simplified and schematized embodiment of a synthetic scheme for polymer sequencing.

FIG. 3 illustrates a coordinate mapping system of a petri plate containing colonies. Each position of a colony can be assigned a distinct coordinate position.

FIG. 4 illustrates various modified embodiments of the substrates.

FIG. 5 illustrates an idealized scanning result corresponding to a particular colony position.

FIG. 6 illustrates particular linkers useful for attaching a nucleic acid to a silicon substrate. Note that thymine may be substituted by adenine, cytidine, guanine, or uracil.

FIG. 7 illustrates an embodiment of the scanning system and reaction chamber.

FIG. 8 illustrates the application of the synthetic scheme for sequencing as applied to a nucleic acid cluster localized to a discrete identified position.

FIG. 8A illustrates schematically, at a molecular level, the sequence of events which occur during a particular sequencing cycle. FIG. 8B illustrates, in a logic flow chart, how the scheme is performed.

FIG. 9 illustrates the synthesis of a representative nucleotide analog useful in the synthetic scheme. Note that the FMOC may be attached to adenine, cytosine, or guanine.

FIG. 10 illustrates the application of the degradative scheme for sequencing as applied to a nucleic acid cluster localized to a discrete identified position. FIG. 10A illustrates schematically, at a molecular level, the sequence of events which occur during a particular sequencing cycle. FIG. 10B illustrates in a logic flow chart how the scheme is performed.

FIG. 11 illustrates a functionalized apparatus for performing the scanning steps and sequencing reaction steps.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

I. Sequencing Procedure for a Generic Polymer

A. Overview

1. Substrate and matrix
2. Scanning system
3. Synthetic/degradative cycles
4. Label
5. Utility

B. Substrate/Matrix

1. Non-distortable
2. Attachment of polymer

C. Scanning system

1. Mapping to distinct position
2. Detection system
3. Digital or analog signal

D. Synthetic or degradative cycle

1. Synthetic cycles
 - a. synthetic scheme
 - b. blocking groups
2. Degradative cycles
3. Conceptual principles

E. Label

1. Attachment
2. Mode of detection

F. Utility

6

II. Specific Embodiments

A. Synthetic method

B. Chain degradation method

III. Apparatus

I. Sequencing Procedure for a Generic Polymer

The present invention provides methods and apparatus for the preparation and use of a substrate having a plurality of polymers with various sequences where each small defined contiguous area defines a small cluster of homogeneous polymer sequences. The invention is described herein primarily with regard to the sequencing of nucleic acids but may be readily adapted to the sequencing of other polymers, typically linear biological macromolecules. Such polymers include, for example, both linear cyclical polymers or nucleic acids, polysaccharides, phospholipids, and peptides having various different amino acids, heteropolymers in which the polymers are mixed, polyurethanes, polyesters, polycarbonates, polyureas, polyamides, polyethyleneimines, polyarylene sulfides, polysiloxanes, polyimides, polyacetates or mixed polymers of various sorts. In a preferred embodiment, the present invention is described in the use of sequencing nucleic acids.

Various aspects of U.S. Ser. No. 07/362,901, (VLSIP® parent); U.S. Ser. No. 07/492,462 (now U.S. Pat. No. 5,143,854) (VLSIP® CIP); U.S. Ser. No. 07/435,316 (caged biotin parent); U.S. Ser. No. 07/612,671 (now U.S. Pat. No. 5,252,743) (caged biotin CIP); and simultaneously filed cases U.S. Ser. No. 07/624,120, a divisional of which has issued as U.S. Pat. No. 5,744,101 (automated VLSIP®); and U.S. Ser. No. 07/624,114, a grandchild of which has issued as U.S. Pat. No. 5,800,992 (sequencing by hybridization); each of which is hereby incorporated by reference, are applicable to the substrates and matrix materials described herein, to the apparatus used for scanning the matrix arrays, to means for automating the scanning process, and to the linkage of polymers to a substrate.

By use of masking technology and photosensitive synthetic subunits, the VLSIPS apparatus allows for the step-wise synthesis of polymers according to a positionally defined matrix pattern. Each oligonucleotide probe will be synthesized at known and defined positional locations on the substrate. This forms a matrix pattern of known relationship between position and specificity of interaction. The VLSIPS technology allows the production of a very large number of different oligonucleotide probes to be simultaneously and automatically synthesized including numbers in excess of about 10^2 , 10^3 , 10^4 , 10^5 , 10^6 , or even more, and at densities of at least about 10^2 , $10^3/\text{cm}^2$, $10^4/\text{cm}^2$, $10^5/\text{cm}^2$ and up to $10^6/\text{cm}^2$ or more. This application discloses methods for synthesizing polymers on a silicon or other suitably derivatized substrate, methods and chemistry for synthesizing specific types of biological polymers on those substrates, apparatus for scanning and detecting whether interaction has occurred at specific locations on the substrate, and various other technologies related to the use of a high density very large scale immobilized polymer substrate. At a size of about 30 microns by 30 microns, one million regions would take about 11 centimeters square or a single wafer of about 4 centimeters by 4 centimeters. Thus the present technology provides for making a single matrix of that size having all one million plus possible oligonucleotides. Region size is sufficiently small to correspond to densities of at least about 5 regions/ cm^2 , 20 regions/ cm^2 , 50 regions/ cm^2 , 100 regions/ cm^2 , and greater, including 300 regions/ cm^2 , 1000 regions/ cm^2 , 3K regions/ cm^2 , 10K regions/ cm^2 , 30K regions/ cm^2 ,

US 7,056,666 B2

7

100K regions/cm², 300K regions/cm² or more, even in excess of one million regions/cm².

A. Overview

The present invention is based, in part, on the ability to perform a step wise series of reactions which either extend or degrade a polymer by defined units.

FIG. 1 schematizes a simplified linear two monomer polymer made up of A type and B type subunits. A degradative scheme is illustrated. Panel A depicts a matrix with two different polymers located at positions 10 and 14, but with no polymer linked at position 12. A reaction is employed to label all of these polymers at the terminus opposite the attachment of the monomer. Panel B illustrates a label (designated by an asterisk) incorporated at position 16 on the terminal monomers. A scan step is performed to locate positions 10 and 14 where polymers have been linked, but no polymer is located at position 12. The entire matrix is exposed to a reagent which is specific for removing single terminal A monomers, which are also labeled. The reagent is selected to remove only a single monomer; it will not remove further A monomers. Removal of the labeled A monomer leaves a substrate as illustrated in panel C. A scan step is performed, and compared with the previous scan, indicates that the polymer located at position 12 has lost its label, i.e., that polymer at 12 terminated with an A monomer. The entire matrix is then exposed to a second reagent which is specific for removing terminal B monomers which are also labeled. Note that only a single B on each monomer is removed and that successive B monomers are not affected. Removal of the labeled B monomer leaves a substrate as illustrated in panel D. Another scan step is performed, indicating that the polymer located at position 14 has lost its label, i.e., it terminated with a B monomer. The sequence of treatments and scans is repeated to determine the successive monomers. It will be recognized that if the labeled A and B are distinguishable, i.e., the label on polymers at sites 10 and 14 may be distinguished, a single removal step can be performed to convert the substrate as illustrated in panel B directly to that illustrated in panel D.

An alternative embodiment employs synthetic reactions where a synthetic product is made at the direction of the attached polymer. The method is useful in the synthesis of a complementary nucleic acid strand by elongation of a primer as directed by the attached polymer.

FIG. 2 illustrates a similar simplified polymer scheme, where the A and B monomer provide a complementary correspondence to A' and B' respectively. Thus, an A monomer directs synthetic addition of an A' monomer and a B monomer directs synthetic addition of a B' monomer. Panel A depicts monomers attached at locations 18 and 22, but not at location 20. Each polymer already has one corresponding complementary monomer A'. The matrix, with polymers, is subjected to an elongation reaction which incorporates, e.g., single labeled A' monomers 24 but not B' monomers, as depicted in panel B. The label is indicated by the asterisk. Note that only one A monomer is added. A scan step is performed to determine whether polymers located at positions 18 or 22 have incorporated the labeled A' monomers. The polymer at position 18 has, while the polymer at position 22 has not. Another elongation reaction which incorporates labeled B' monomers 26 is performed resulting in a matrix as depicted in panel C. Again note that only one, and not successive B' monomers, is added. Another scan is performed to determine whether a polymer located at sites 18 or 22 has incorporated a labeled B' monomer, and the result indicates that the polymer located at site 22 has incorporated the labeled B' monomer. A next step removes

8

all of the labels to provide a substrate as depicted in panel D. As before, if the polymer which incorporated a labeled A' monomer is distinguishable from a polymer which incorporated a labeled B' monomer, the separate elongation reactions may be combined producing a panel C type matrix directly from a panel A type matrix and the scan procedure can distinguish which terminal monomer was incorporated.

It will be appreciated that the process may be applied to more complicated polymers having more different types of monomers. Also, the number of scan steps can be minimized if the various possible labeled monomers can be differentiated by the detector system.

Typically, the units will be single monomers, though under certain circumstances the units may comprise dimers, trimers, or longer segments of defined length. In fact, under certain circumstances, the method may be operable in removing or adding different sized units so long as the units are distinguishable. However, it is very important that the reagents used do not remove or add successive monomers. This is achieved in the degradative method by use of highly specific reagents. In the synthetic mode, this is often achieved with removable blocking groups which prevent further elongation.

One important aspect of the invention is the concept of using a substrate having homogeneous clusters of polymers attached at distinct matrix positions. The term "cluster" refers to a localized group of substantially homogeneous polymers which are positionally defined as corresponding to a single sequence. For example, a coordinate system will allow the reproducible identification and correlation of data corresponding to distinct homogeneous clusters of polymer locally attached to a matrix surface. FIG. 3 illustrates a mapping system providing such a correspondence, where transfer of polymers from a producing colony of organisms to a matrix preserves spatial information thereby allowing positional identification. The positional identification allows correlation of data from successive scan steps.

In one embodiment, bacterial colonies producing polymers are spatially separated on the media surface of a petri plate as depicted in panel A. Alternatively, phage plaques on a bacterial lawn can exhibit a similar distribution. A portion of panel A is enlarged and shown in panel B. Individual colonies are labeled C1-C7. The position of each colony can be mapped to positions on a coordinate system, as depicted in panel C. The positions of each colony can then be defined, as in a table shown in panel D, which allows reproducible correlation of scan cycle results.

Although the preferred embodiments are described with respect to a flat matrix, the invention may also be applied using the means for correlating detection results from multiple samples after passage through batch or continuous flow reactions. For example, spatially separated polymers may be held in separate wells on a microtiter plate. The polymers will be attached to a substrate to retain the polymers as the sequencing reagents are applied and removed.

The entire substrate surface, with homogeneous clusters of polymer attached at defined positions, may be subjected to batch reactions so the entire surface is exposed to a uniform and defined sequence of reactions. As a result, each cluster of target polymers for sequencing will be subjected to similar reactive chemistry. By monitoring the results of these reactions on each cluster localized to a defined coordinate position, the sequence of the polymer which is attached at that site will be determined.

FIG. 4, panel A illustrates solid phase attached polymers linked to particles 32 which are individually sequestered in separate wells 34 on a microtiter plate. The scanning system

US 7,056,666 B2

9

will separately scan each well. FIG. 4 panel B illustrates marbles 36 to which polymers are attached. The marbles are automatically fed in a continuous stream through the reaction reagents 38 and past a detector 40. The marbles may be carefully held in tubes or troughs which prevent the order of the beads from being disturbed. In a combination of the two embodiments, each polymer is attached to a plurality of small marbles, and marbles having each polymer are separated, but retained in a known order. Each marble is, in batch with a number of analogous marbles having other polymers linked individually to them, passed through a series of reagents in the sequencing system. For example, A2, B2, and C2 are subjected to sequencing reactions in batch, with label incorporated only for the second monomer A3, B3, and C3, are likewise treated to determine the third monomer. Likewise for A_n, B_n, and C_n. However, within each batch, the detection will usually occur in the order A, B, and C, thereby providing for correlation of successive detection steps for the A polymer beads, for the B polymer beads, and for the C polymer beads.

FIG. 5 illustrates a signal which might result from a particular defined position. Panel A illustrates the position of a given colony relative to the positions corresponding to the positional map. The scan system will typically determine the amount of signal, or type of signal, at each position of the matrix. The scan system will adjust the relationship of the detector and the substrate to scan the matrix in a controllable fashion. An optical system with mirrors or other elements may allow the relative positions of the substrate and detection to be fixed. The scanner can be programmed to scan the entire substrate surface in a reproducible manner, or to scan only those positions where polymer clusters have been localized. A digital data map, panel B, can be generated from the scan step.

Thus, instead of subjecting each individual and separated polymer to the series of reactions as a homogeneous sample, a whole matrix array of different polymers targeted for sequencing may be exposed to a series of chemical manipulations in a batch format. A large array of hundreds, thousands, or even millions of spatially separated homogeneous regions may be simultaneously treated by defined sequencing chemistry.

The use of a coordinate system which can reproducibly assay a defined position after each reaction cycle can be advantageously applied according to this invention. For example, a colony plaque lift of polymers can be transferred onto a nitrocellulose filter or other substrate. A scanning detector system will be able to reproducibly monitor the results of chemical reactions performed on the target polymers located at the defined locations of particular clones. An accurate positioning can be further ensured by incorporating various alignment marks on the substrate.

The use of a high resolution system for monitoring the results of successive sequencing steps provides the possibility for correlating the scan results of each successive sequencing reaction at each defined position.

The invention is dependent, in part, upon the stepwise synthesis or degradation of the localized polymers as schematized in FIGS. 1 and 2. The synthetic scheme is particularly useful on nucleic acids which can be synthesized from a complementary strand. Otherwise, a stepwise degradation scheme may be the preferred method. Although single monomer cycles of synthesis or degradation will usually be applicable, in certain cases the technology will be workable using larger segments, e.g., dimers or trimers, in the cyclic reactions.

10

The present invention also provides methods for production or selection of monomer-specific degradative reagents based upon catalytic antibody constructs. Antibody binding sites exhibiting specificity for binding particular terminal monomers can be linked to cleavage reagents or active sites of cleavage enzymes. Thus, reagents which are specific for particular terminal nucleotides may function to remove them in a specific fashion.

The invention also makes use of a means for detecting or labeling the polymers. Particular sequencing chemistry can be selected for specificity in reacting with terminal monomer units. Alternatively, indirect labeling methods may be applied which can distinguish between different terminal monomers. Another alternative scheme allows for terminal labeling which is not monomer-specific, but with the determination of the monomer based upon specificity of post-label reagents or upon monomer-distinguishable labels. Suitable such reagents will be antibodies or other reagents having specificity for distinguishing between different labeled terminal monomer residues and cleaving only those labeled monomer residues.

Thus, although neither the reaction nor the label need necessarily be specific, at least one of the pair must be specific. This ensures that a comparison of label signal before and after a reaction allows determination of the change in label signal after monomer specific reactions are performed, and thereby providing the means to deduce the identity of the monomer at a given position.

B. Substrate/Matrix

The substrate or matrix has relatively few constraints on its composition. Preferably, the matrix will be inert to the sequencing reactions to which the polymers attached thereto will be subjected. Typically, a silicon or glass substrate will be used, but other suitable matrix materials include ceramics, or plastics, e.g., polycarbonate, polystyrene, delrin, and cellulose, and any other matrix which satisfies these functional constraints.

In one embodiment, the matrix should be sufficiently nondeformable that the scanning system can reproducibly scan the matrix and reliably correlate defined positions with earlier and later scan operations. However, by including alignment markings on the substrate, the need for absolute rigidity of the substrate may be reduced.

In an alternative embodiment, the matrix may merely be large enough that the attached polymer may be separated from a liquid phase containing the sequencing reagents. In this embodiment, a single detection unit is used to analyze the label in a multiplicity of different samples after each of the reaction steps. Thus, different samples may be separately treated in distinct wells of a microtiter dish.

Separate homogeneous polymers can be introduced to solid phase beads in each microtiter well. Sequencing reagents may be individually introduced separately into each well, or transferred from well to well with the polymers remaining in the correct well due to their solid phase attachments.

In an alternative approach, the solid phase matrix may be marbles or other particularly shaped articles. Spherical shapes, solid or hollow, are preferred because they can be easily transported through troughs or tubing which retains their relative orders. By feeding a succession of beads through appropriate reaction baths and past a detector in a known and retained order, a succession of label detection results from a bead may be correlated and converted into a polymer sequence.

The attachment of the target homogeneous clusters of target polymers to the substrate can be achieved by appro-

US 7,056,666 B2

11

appropriate linkage chemistry. As indicated before, the linkage should be stable and insensitive to the sequencing reagents used. The specific linkages will depend, of course, upon the particular combination of substrate and polymer being used.

Typically, the most useful chemical moieties which will be used are amines. Typical substrate derivatized groups include aminopropyl triethoxysilane, hydroxypropylacrylate, or hydroxy reagents, see, e.g., U.S. Ser. No. 07/624,120, a divisional of which has issued as U.S. Pat. No. 5,744,101 (automated VLSIPS®). Typical polymer derivatized groups include nitroveratryl and nitroveratryl oxycarbonyl. Linkage types are also illustrated and detailed in U.S. Ser. No. 07/624,120, a divisional of which has issued as U.S. Pat. No. 5,744,101 (automated VLSIPS®), and U.S. Ser. No. 07/624,114, (a grandchild of which has issued as U.S. Pat. No. 5,800,992) (sequencing by hybridization).

FIG. 6 illustrates one preferred linkage chemistry for nucleic acids. An NVO-derivatized U.S. Ser. No. 07/624,120, a divisional of which has issued as U.S. Pat. No. 5,744,101 (automated VLSIPS®). The specific conditions for synthesis of thymidine are described therein and are adaptable to other nucleotides and nucleosides. The nucleoside analog is further derivatized with an appropriate R group at the 3' hydroxyl. Preferred R groups are indicated in FIG. 6. The linkage produces a photosensitive blocked nucleoside suitable for phosphoramidite synthesis of further polynucleotides which can serve as a complementary strand for hybridization of other polymers. The hybrids of the complementary strands may be covalently crosslinked using acridine dyes or other intercalative reagents, e.g., psoralen. See, e.g., Kornberg (1980) *DNA Replication* Freeman, San Francisco; Wiesehahn, et al. (1978) *Proc. Natl. Acad. Sci. USA* 75:2703-2707, and Sheldon (1986) U.S. Pat. No. 4,582,789 which are each incorporated herein by reference.

The linkage should be substantially inert to the cyclic sequencing reactions and scan cycles. Usually, the linkage will be at a defined and homogeneous polymer position, preferably at the end opposite where the sequencing chemistry takes place. Although the type of linkage is dependent upon the polymer being sequenced, various types of polymers have preferred linkages. For polypeptides, amino terminal or carboxyl terminal linkages will be preferred. Specific amino terminal linkages include amino butyric acid, amino caproic acids, and similar carboxylic acids. Specific carboxyl terminal linkages include butyric acid, caproic acid, and other carboxylic acids, hydrocarbon, and ethers. See U.S. Ser. No. 07/435,316 (VLSIPS® parent), and U.S. Ser. No. 07/492,462 (VLSIP® CIP) (now U.S. Pat. No. 5,143,854), which are incorporated herein by reference. For nucleic acids, the linkages will typically be either 5' or 3' linkages. Suitable 3' linkages include those illustrated in FIG. 6, and others described in U.S. Ser. No. 07/624,114, (a grandchild of which has issued as U.S. Pat. No. 5,800,992) (sequencing by hybridization).

Alternatively, for complementary polymers, particularly nucleic acids, linkage may be via crosslinkage of the complementary polymers where the complementary strand is directly attached to the matrix. Acridine dyes, e.g., psoralen, or a similar crosslinking agent between the strands can be used. See, e.g., Dattagupta, et al., "Coupling of Nucleic Acids to Solid Support By Photochemical Methods," U.S. Pat. No. 4,713,326; and U.S. Pat. No. 4,542,102; and Chatterjee, M. et al. (1990) *J. Am. Chem. Soc.* 112:6397; which describe useful crosslinking reagents, and are hereby incorporated herein by reference.

For polynucleotides, the preferred attachment to the matrix is through a synthetic oligomer by the 5' end of each

12

target sequence. This oligomer is designed to anneal to the desired target templates used in a synthetic system or to the polynucleotide used in the degradation approach. In one embodiment, a vector sequence which is complementary to the immobilized oligonucleotide is incorporated adjacent the cloning inserts, thereby providing a common complementary sequence for each insert. In particular, a cloning vector will be selected with a defined sequence adjacent the insert. See, e.g., Sambrook et al., (1989) *Molecular Cloning: A Laboratory Manual*, Vols. 1-3, Cold Spring Harbor Press, which is hereby incorporated herein by reference. This defined sequence is used, in some embodiments, as a common linker for all of the vector inserts. The inserts, adjacent to this linker, will be transferable by hybridization to the matrix linked complementary sequences. The hybrids are crosslinked by addition of a suitable crosslinker under appropriate conditions, for example, photocrosslinking by psoralen with uv light. See, e.g., Song et al. (1979) *Photochem. Photobiol.* 29:1177-1197; Cimino et al. (1985) *Ann. Rev. Biochem.* 54:1151-1193; and Parsons (1980) *Photochem. Photobiol.* 32:813-821; each of which is incorporated herein by reference. Using these approaches, the oligonucleotide linker serves as both the attachment linker and the polymerization primer.

FIG. 6 illustrates a preferred 3' terminal linkage designed for a phosphoramidite linkage of a synthetic primer and the reactions forming them. The chemical reactions for actually performing the linkage will be similar to those used for oligonucleotide synthesis instruments using phosphoramidite or similar chemistry. Applied Biosystems, Foster City, Calif. supplies oligonucleotide synthesizers.

C. Scanning System

The scanning system should be able to reproducibly scan the substrate. Where appropriate, e.g., for a two dimensional substrate where the polymers are localized to positions thereon, the scanning system should positionally define the clusters attached thereon to a reproducible coordinate system. It is important that the positional identification of clusters be repeatable in successive scan steps. Functionally, the system should be able to define physical positions to a coordinate system as described above and illustrated in FIGS. 3 and 4.

In alternative embodiments, the system can operate on a cruder level by separately detecting separate wells on a microtiter plate, or by scanning marbles which pass by the detector in an embodiment as described above and illustrated in FIG. 4.

The scanning system would be similar to those used in electrooptical scanning devices. See, e.g., the fluorescent detection device described in U.S. Ser. No. 07/492,462 (VLSIP® CIP), now U.S. Pat. No. 5,143,854, and U.S. Ser. No. 07/624,120, a divisional of which has issued as U.S. Pat. No. 5,744,101 (automated VLSIPS®). The system could exhibit many of the features of photographic scanners, digitizers or even compact disk reading devices. For example, a model no. PM500-A1 x-y translation table manufactured by Newport Corporation can be attached to a detector unit. The x-y translation table is connected to and controlled by an appropriately programmed digital computer such as an IBM PC/AT or AT compatible computer. The detection system can be a model no. R943-02 photomultiplier tube manufactured by Hamamatsu, attached to a preamplifier, e.g., a model no. SR440 manufactured by Stanford Research Systems, and to a photon counter, e.g., an SR430 manufactured by Stanford Research System, or a multichannel detection device. Although a digital signal may

US 7,056,666 B2

13

usually be preferred, there may be circumstances where analog signals would be advantageous.

The stability and reproducibility of the positional localization in scanning will determine, to a large extent, the resolution for separating closely positioned polymer clusters in a 2 dimensional substrate embodiment. Since the successive monitoring at a given position depends upon the ability to map the results of a reaction cycle to its effect on a positionally mapped cluster of polymers, high resolution scanning is preferred. As the resolution increases, the upper limit to the number of possible polymers which may be sequenced on a single matrix will also increase. Crude scanning systems may resolve only on the order of 1000 μ , refined scanning systems may resolve on the order of 100 μ , more refined systems may resolve on the order of about 10 μ with optical magnification systems a resolution on the order of 1.0 μ is available, and more preferably a resolution on the order of better than 0.01 μ is desired. The limitations on the resolution may be diffraction limited and advantages may arise from using shorter wavelength radiation for the photo-optical deprotection fluorescent scanning steps. However, with increased resolution, the time required to fully scan a matrix will be increased and a compromise between speed and resolution will necessarily be selected. Parallel detection devices which will provide high resolution with shorter scan times will be applicable where multiple detectors will be moved in parallel.

With other embodiments, resolution often is not so important and sensitivity might be emphasized. However, the reliability of a signal may be pre-selected by counting photons and continuing to count for a longer period at positions where intensity of signal is lower. Although this will decrease scan speed, it can increase reliability of the signal determination. Various signal detection and processing algorithms may be incorporated into the detection system, such as described in U.S. Ser. No. 07/624,120, a divisional of which has issued as U.S. Pat. No. 5,744,101 (activated VLSIPS). In one embodiment, the distribution of signal intensities of pixels across the region of signal are evaluated to determine whether the distribution of intensities corresponds to a time positive signal.

The detection system for the signal or label will depend upon the label used, which may be defined by the chemistry available. For optical signals, a combination of an optical fiber or charged couple device (CCD) may be used in the detection step. In those circumstances where the matrix is itself transparent to the radiation used, it is possible to have an incident light beam pass through the substrate with the detector located opposite the substrate from the polymers. For electromagnetic labels, various forms of spectroscopy systems can be used. Various physical orientations for the detection system are available and discussion of important design parameters is provided, e.g., in Jovin, Adv. in Biochem. Biophys., which is hereby incorporated herein by reference.

Various labels which are easily detected include radioactive labels, heavy metals, optically detectable labels, spectroscopic labels and the like. Various photoluminescent labels include those described in U.S. Ser. No. 07/624,114, a grandchild of which has issued as U.S. Pat. No. 5,800,992 (sequencing by hybridization). Protection and deprotection are described, e.g., in McCray, et al. (1989) *Ann. Rev. Biophysical Chemistry* 18:239-270, and U.S. Ser. No. 07/624,120, a divisional of which has issued as U.S. Pat. No. 5,744,101 (automated VLSIP®), each of which is hereby incorporated herein by reference.

14

With a processing system, the speed of scanning may be dramatically increased with a system which only scans positions where known clusters of polymer are attached. This allows the scanning mechanism to skip over areas which have been determined to lack any polymer clusters and avoids loss of time in scanning useless regions of the matrix. Moreover, various problems with spurious or overlapping signals may be adjusted for by appropriate analysis.

A scanning apparatus which may be used for the presently described uses is schematically illustrated in FIG. 7. A substrate 52 is placed on an x-y translation table 54. In a preferred embodiment the x-y translation table is a model no. PM500-A1 manufactured by Newport Corporation. The x-y translation table is connected to and controlled by an appropriately programmed digital computer 56 which may be, for example, an appropriately programmed IBM PC/AT or AT compatible computer. Of course, other computer systems, special purpose hardware, or the like could readily be substituted for the AT computer used herein for illustration. Computer software for the translation and data collection functions described herein can be provided based on commercially available software including, for example, "Lab Windows" licensed by National Instruments, which is incorporated herein by reference for all purposes.

The substrate and x-y translation table are placed under a microscope 58 which includes one or more objectives 60. Light (about 488 nm) from a laser 62, which in some embodiments is a model no. 2020-05 argon ion laser manufactured by Spectraphysics, is directed at the substrate by a dichroic mirror 64 which passes greater than about 520 nm wavelength light but reflects 488 nm light. Dichroic mirror 64 may be, for example, a model no. FT510 manufactured by Carl Zeiss. Light reflected from the mirror then enters the microscope 58 which may be, for example, a model no. Axioscop 20 manufactured by Carl Zeiss. Fluorescein-marked materials on the substrate will fluoresce >488 nm light, and the fluoresced light will be collected by the microscope and passed through the mirror. The fluorescent light from the substrate is then directed through a wavelength filter 66 and, thereafter through an aperture plate 68. Wavelength filter 66 may be, for example, a model no. OG530 manufactured by Melles Griot and aperture plate 68 may be, for example, a model no. 477352/477380 manufactured by Carl Zeiss.

The fluoresced light then enters a photomultiplier tube 70 which in one embodiment is a model no. R943-02 manufactured by Hamamatsu, the signal is amplified in preamplifier 72 and photons are counted by photon counter 74. The number of photons is recorded as a function of the location in the computer 56. Pre-Amp 72 may be, for example, a model no. SR440 manufactured by Stanford Research Systems and photon counter 74 may be a model no. SR430 manufactured by Stanford Research Systems. The substrate is then moved to a subsequent location and the process is repeated. In preferred embodiments the data are acquired every 1 to 100 μ m with a data collection diameter of about 0.8 to 10 μ m preferred. In embodiments with sufficiently high fluorescence, a CCD detector with broadfield illumination is utilized.

By counting the number of photons generated in a given area in response to the laser, it is possible to determine where fluorescent marked molecules are located on the substrate. Consequently, for a substrate which has a matrix of polypeptides, for example, synthesized on the surface thereof, it is possible to determine which of the polypeptides has incorporated a fluorescently marked monomer.

US 7,056,666 B2

15

According to preferred embodiments, the intensity and duration of the light applied to the substrate is controlled by varying the laser power and scan stage rate for improved signal-to-noise ratio by maximizing fluorescence emission and minimizing background noise. Signal analysis may improve the resolution and reliability of the system. The time of photon counting may be varied at various positions to provide high signal to background or noise.

D. Synthetic or Degradative Cycle

The present invention provides a substrate with positionally separated polymers for sequencing. The separation may be by solid phase carriers separated in separate wells, by separately manipulable carriers such as beads or marbles, or by physical separation of regions on a two-dimensional substrate surface. Each cluster region is a target for the sequencing reactions. Although the reactions are, in various embodiments, performed on all the clusters together, each cluster can be individually analyzed by following the results from the sequence of reactions on polymer clusters at positionally defined locations.

The synthetic mode, as illustrated in FIG. 1 is easily applied to the sequencing of nucleic acids, since one target strand may serve as the template to synthesize the complementary strand. The nucleic acid can be DNA, RNA or mixed polymers. For the purposes of illustration, and not by limitation, the sequencing steps for DNA are described in detail. The synthetic mode, an example of which is depicted in FIG. 8 for nucleotides, may also be useful in circumstances where synthesis occurs in response to a known polymer sequence. The synthetic scheme depends, in part, on the stepwise elongation by small and identifiable units. A polymerase is used to extend a primer complementary to a target template. The primer is elongated one nucleotide at a time by use of a particular modified nucleotide analog to which a blocking agent is added and which prevents further elongation. This blocking agent is analogous to the dideoxy nucleotides used in the Sanger and Coulson sequencing procedure, but in certain embodiments here, the blockage is reversible. This analog is also labeled with a removable moiety, e.g., a fluorescent label, so that the scanning system can detect the particular nucleotide incorporated after its addition to the polymerization primer.

Panel 4A illustrates the cycle of sequence reactions in one embodiment. The template polymer **82** located at a particular site has already been linked to substrate. The template **82** and complementary primer **84** are hybridized. Often, the primer **84** is common to all of the target template sequences, selected by its common occurrence on a selected cloning vector. The primer **84** is also often covalently crosslinked to the target template **82** using psoralen and U.V. light.

Labeled and blocked monomers **86** are shown, the label depicted by the asterisk and the polymerization blocking groups indicated by B. A compatible polymerase **88** which can elongate the primer with the labeled blocked monomers **86** is used in reaction 1. In the preferred embodiment, the separate labeled monomers can be distinguished from one another by the wavelength of fluorescent emission.

In the example illustrated, a labeled blocked guanosine monomer has been incorporated into the elongated primer **90**.

Step 2 is a scan, where the signal at the position corresponding to template **82** indicates that the guanosine analog was incorporated. Reaction 2 is performed, a removal of both the label and blocking group. It will be recognized that the blocking group prevents elongation by any more than a single nucleotide in each reaction cycle. Reaction 3 is

16

equivalent to reaction 1, though the substrate primer has been elongated by one monomer.

Panel B illustrates the scheme in a logic flow chart. The template **82** is attached to the substrate, either directly or through the primer. Reaction 1 elongates the primer by a single labeled blocked nucleotide. A scan step is performed and the blocking and labeling agents are removed. The elongation reaction is performed and the cycle repeated.

For a nucleic acid, a unit for addition would typically be a single nucleotide. Under certain circumstances, dimers or trimers or larger segments may be utilized, but a larger number of different possible nucleotide elements requires high distinguishability in other steps. For example, there are only four different nucleotide monomer possibilities, but there are sixteen different dimer possibilities. The distinction among four possibilities is more precise and simple than among sixteen dimer possibilities. To prevent elongation by a unit length greater than one monomer, the nucleotide should be blocked at the position of 3' elongation. Usually, the nucleotide will be blocked at the 3' hydroxyl group where successive nucleotides would be attached. In contrast to a dideoxy nucleotide, typically the blocking agent will be a reversible blocking agent thereby allowing for deblocking and subsequent elongation.

Variations may be easily incorporated into the procedure. If the labels on the monomers are not distinguishable, successive substrate scans can be performed after each monomer is provided conditions allowing its incorporation. Alternatively, a small fraction of permanently blocked but reversibly labeled monomers may be incorporated. Those specific molecules which incorporate the blocked monomers are permanently removed from further polymerization, but such is acceptable if the labeling moiety is also removed.

1. Other Monomers

one important functional property of the monomers is that the label be removable. The removal reaction will preferably be achieved using mild conditions. Blocking groups sensitive to mild acidic conditions, mild basic conditions, or light are preferred. The label position may be anywhere on the molecule compatible with appropriate polymerization, i.e., complementary to the template, by the selected polymerase. A single polymerase for all of the modified nucleotide is preferred, but a different polymerase for each of the different monomers can be used.

Nucleotide analogs used as chain-terminating reagents will typically have both a labeling moiety and a blocking agent while remaining compatible with the elongation enzymology. As the blocking agent will usually be on the 3' hydroxyl position of the sugar on a nucleotide, it would be most convenient to incorporate the label and the blocking agent at the same site, providing for a single reaction for simultaneous removal of the label and blocking agent. However, it is also possible to put a label on another portion of the nucleotide analog than the 3' hydroxyl position of the sugar, thereby requiring a two-step reaction cycle for removing the blocking and labeling groups.

Analogous will be found by selecting for suitable combinations of appropriate nucleotides with compatible polymerases. In particular, it is desired that a selected polymerase be capable of incorporating a nucleotide, with selectivity, having both the blocking moiety and the label moiety attached. It has been observed that RNA polymerases are less fastidious with respect to the nucleotide analogues which will be polymerized into a growing chain. See, e.g., Rozovaskaya, T., et al. (1977) *Molekulyarnaya Biologiya*, 11:598-610; Kutateladze, T., et al. (1986) *Molekulyarnaya Biologiya*, 20:267-276; and Chidgevadze, Z., et al. (1985)

US 7,056,666 B2

17

FEBS Letters, 183:275–278. Moreover, those references also indicate that rather significant chemical moieties may be attached at the 2' or 3' positions on a nucleotide, and still be correctly incorporated at the growing chain terminus.

In particular, it is not necessary that the same nucleotide have both the reversible blocking moiety and the removable labeling moiety, as a combination of two separate nucleotide analogues could be utilized, e.g., N1, which is reversibly blocked and not labeled, and N2, which is irreversibly blocked but removably labeled. Note that the removal of label may be affected by destruction of the label, e.g., fluorescence destruction, or preferably by removal. Both of these nucleotides might be, for instance, A analogues. With the mixture, at an appropriate sequence position of a target sequence, an N1 and N2 nucleotides can be incorporated at an appropriate ratio, and these can be polymerized by either two separate polymerases, or preferably a single polymerase.

For example, two separate polymerases might be necessary, P1 which incorporates N1, and P2 which incorporates N2. At the given location in the sequence, some of the growing polymers will incorporate N1 with P1 polymerase, and others will incorporate N2 with the P2 polymerase. The proportions of N1, N2, P1, and P2 may be titrated to get the desired fractional proportions of the N1 reversibly blocked nucleotides and the N2 labeled but irreversibly blocked nucleotides.

As all of the growing chains have blocked nucleotides, no elongation takes place beyond a single nucleotide. The N2 nucleotides provide a specific label, detected in the scanning step. After determination of the incorporated label, the label may be removed or destroyed, and those irreversibly terminated growing chains become permanently removed from further participation in the sequencing process. Photodestruction may be achieved by a high intensity laser beam of the correct wavelength. See, e.g., March (1977) *Advanced Organic Chemistry: Reactions, Mechanisms and Structure* (2d Ed) McGraw; and Carey and Sundberg (1980) *Advanced Organic Chemistry: part A Structure and Mechanisms*, Plenum.

Next, the reversible blocking moiety is removed, providing a new set of slightly longer polymers ready for the next step. Of course, the amount of label necessary to be incorporated must be detectable, preferably with a clear, unambiguous positive signal. The amount of label incorporated will depend, in part, upon the conditions in the polymerizing step and the relative incorporation of the N1 and N2 nucleotides. The proportions of the nucleotides, polymerases, and other reagents may be adjusted to appropriately incorporate the desired proportions of the nucleotides.

In an embodiment where a single polymerase will incorporate both N1 and N2, the relative proportions and conditions to get the correct incorporation levels of the two nucleotides can be titrated. In an alternative preferred embodiment, a single nucleotide will have both the removable label and the reversible blocking moiety.

A similar approach may be necessary where only some fraction of the nucleotide analogues is labeled. Separate polymerases might also be useful for such situations, and each polymerase may have special conditions necessary for activity.

Procedures for selecting suitable nucleotide and polymerase combinations will be readily adapted from Ruth et al. (1981) *Molecular Pharmacology* 20:415–422; Kutateladze, T., et al. (1984) *Nuc. Acids Res.*, 12:1671–1686; Kutateladze, T., et al. (1986) *Molekulyarnaya Biologiya* 20:267–276;

18

Chidgeavadze, Z., et al. (1985) *FEBS Letters*, 183:275–278; and Rozovskaya, T., et al. (1977) *Molekulyarnaya Biologiya* 11:598–610.

The determination of termination activity is done in two steps. First, nucleotide analogues are screened for the ability of the compound to inhibit polymerase activity. Then the nucleotide analogue is tested for base-specific termination as manifested by generating a correct DNA sequencing ladder on a template of known sequence. The appropriate reaction conditions are those used for conventional sequencing reactions with the respective polymerases. The conditions are then modified in the usual ways to obtain the optimal conditions for the particular terminator compound (e.g. concentration of terminator, ratio of terminator to dNTP, Mggt, and other reagents critical to proper polymerase function.

By way of example, an approach employing the polymerase known as reverse transcriptase (AMV) will be described. The initial conditions are essentially as described by Prober, et al. (1987) *Science* 238: 336–341.

A nucleotide analogue is first selected from the group available from a commercial source such as Amersham, New England Nuclear, or Sigma Chemical Company. In particular, nucleotides which are reversibly blocked from further elongation, especially at the 5' or 3' —OH will be used.

General properties which are desired have been described. Each of these analogs can be tested for compatibility with a particular polymerase by testing whether such polymerase is capable of incorporating the labeled analog. Various polymerases may be screened, either natural forms of the mentioned types, or variants thereof. Polymerases useful in connection with the invention include *E. Coli* DNA polymerase (Klenow fragment); and Klenow and Henningsen (1970) *Proc. Nat'l Acad Sci USA* 65:168–175; and Jacobsen et al. (1974) *Eur. J. Biochem.* 45:623–627; modified and cloned versions of T7 DNA polymerase (Sequenase™ and Sequenase 2.0™); see Tabor and Richardson (1987) *Proc. Nat'l Acad. Sci. USA* 84:4767–4771; and Tabor and Richardson (1987) *J. Biol. Chem.* 262:15330–15333; Taq DNA polymerase from thermostable *Thermus aquaticus*; see Chien et al. (1976) *J. Bacteriol.* 127:1550–1557; and its cloned version Amplitaq™; Saiki and Gelfand (1989) *Amplifications* 1:4–6; T4 DNA polymerase; see Nossal (1974) *J. Biol. Chem.* 249:5668–5676, and various reverse transcriptases, both RNA- and DNA-dependent DNA polymerases, e.g., avian retroviruses; see Houts (1979) *J. Virology* 29:517–522; and murine retroviruses; see Kotewicz et al. (1985) *Gene* 35:249–258; Gerard et al. (1986) *DNA* 5:271–279 and Bst polymerase; see Ye, S. and Hong (1987) *Scientia Sinica* 30:503–506.

In order to ensure that only a single nucleotide is added at a time, a blocking agent is usually incorporated onto the 3' hydroxyl group of the nucleotide. Optimally, the blocking agent should be removable under mild conditions (e.g., photosensitive, weak acid labile, or weak base labile groups), thereby allowing for further elongation of the primer strand with a next synthetic cycle. If the blocking agent also contains the fluorescent label, the dual blocking and labeling functions will be achieved without the need for separate reactions for the separate moieties.

The blocking group should have the functional properties of blocking further elongation of the polymer. Additional desired properties are reversibility and inertness to the sequencing reactions. Preferably, where an enzymatic elongation step is used, the monomers should be compatible with the selected polymerase. Specific examples for blocking

US 7,056,666 B2

19

groups for the nucleic acids include acid or base labile groups at the 3'OH position. See, e.g., Gait (1984) *Oligonucleotide Synthesis: A Practical Approach*, IRL Press, Oxford.

A DNA-dependent DNA polymerase is the polymerases of choice. Polymerases used for conventional DNA sequencing, for example, Klenow fragment of *E. coli* DNA Pol, Sequenase (modified T7 DNA polymerase), Taq (*Thermus aquaticus*) DNA polymerase, Bst (*Bacillus stearothermophilus*), DNA polymerase, reverse transcriptase (from AMV, MMLV, RSV, etc.) or other DNA polymerases will be the polymerases of choice. However, there is a functional constraint that the polymerase be compatible with the monomer analogues selected. Screening will be performed to determine appropriate polymerase and monomer analog combinations.

Removal of the blocking groups may also be unnecessary if the labels are removable. In this approach, the chains incorporating the blocked monomers are permanently terminated and will no longer participate in the elongation processes. So long as these blocked monomers are also removed from the labeling process, a small percentage of permanent loss in each cycle can also be tolerated.

The fluorescent label may be selected from any of a number of different moieties. The preferred moiety will be a fluorescent group for which detection is quite sensitive. Various different fluorescence-labeling techniques are described, for example, in Cambara et al. (1988) "Optimization of Parameters in a DNA Sequenator Using Fluorescence Detection," *Bio/Technol.* 6:816-821; Smith et al. (1985) *Nucl. Acids Res.* 13:2399-2412; and Smith et al. (1986) *Nature* 321:674-679, each of which is hereby incorporated herein by reference. Fluorescent labels exhibiting particularly high coefficients of destruction may also be useful in destroying nonspecific background signals.

Appropriate blocking agents include, among others, light sensitive groups such as 6-nitroveratryl-oxycarbonyl (NVOC), 2-nitrobenzyloxycarbonyl (NBOC), α,α -dimethyldimethoxybenzyloxycarbonyl (DDZ), 5-bromo-7-nitroindolyl, o-hydroxy-2-methyl cinnamoyl, 2-oxymethylene anthraquinone, and t-butyl oxycarbonyl (TBOC). Other blocking reagents are discussed, e.g., in U.S. Ser. No. 07/492,462; Patchornik (1970) *J. Amer. Chem. Soc.* 92:6333; and Amit et al. (1974) *J. Org. Chem.* 39:192, all of which are hereby incorporated herein by reference. Additional blocking agents attached to particular positions may be selected according to the functional directives provided herein.

FIG. 9 schematically illustrates the synthesis of a generic protected nucleotide. A suitable nucleotide is labeled with the FMOC fluorescently detectable label by reaction under the conditions described, e.g., in U.S. Ser. No. 07/624,114, (a grandchild of which has issued as U.S. Pat. No. 5,800,992) (sequencing by hybridization), with (TMS-Cl), FMOC-Cl, and H₂O. A protection moiety will be added using conditions also described there.

Various nucleotides possessing features useful in the described method can be readily synthesized. Labeling moieties are attached at appropriate sites on the nucleotide using chemistry and conditions as described, e.g., in Gait (1984) *Oligonucleotide Synthesis*. Blocking groups will also be added using conditions as described, e.g., in U.S. Ser. No. 07/624,114, (a grandchild of which has issued as U.S. Pat. No. 5,800,992) (sequencing by hybridization). FIG. 9 also outlines various reactions which lead to useful nucleotides.

Additionally, the selected polymerases used in elongation reactions should be compatible with nucleotide analogs

20

intended for polymerization to the primer. Simple screening procedures for nucleotide and polymerase combinations may be devised to verify that a particular combination is functional. A test using primer with template which directs the addition of the nucleotide analog to be incorporated will determine whether the combination is workable. Natural polymerases or variants thereof may be used under particular defined conditions.

The degradative scheme is generally illustrated in FIG. 1, an example more generally applicable to biological macromolecular polymers is depicted in FIG. 10. This method is useful for a wider variety of polymers without the limitations imposed by the need to replicate the polymer. The degradative sequencing technique depends, in part, upon the ability to specifically label or distinguish between various different terminal monomers at particular matrix positions. Reactions for specific removal of a defined monomer unit are important.

This monomer distinguishability can arise from an ability to differentiate between label on the various possible monomers in the polymer. As a second means, distinguishability can come from specific reagents which react with particularity on different monomers. Thus, for instance, labels may be used which generally attach to the terminal nucleotide, but whose fluorescent signal differs depending upon the nucleotide. As a third means, a reagent which specifically affects the label on only one monomer may be used, as described below.

In the first example, every polymer cluster will be labeled at a particular end, e.g., the 5' end, without specificity for the monomer located there. The scan step will be able to distinguish the terminal monomers, after which each labeled terminal monomer is specifically removed. The general label step is repeated in the cycle as described.

In the second means for distinguishability, reagents are used which produce a signal which is dependent upon the terminal nucleotide. For example, a labeling molecule which binds only to one specific terminal monomer will provide a monomer specific label. This will provide a cycle much like the first means for distinguishability where the properties of the label is different depending upon the terminal nucleotide to which each specific labeling reagent binds.

In the third means for distinguishability, an individual reagent labels or affects only a specific terminal monomer. Polymers susceptible to each reagent by virtue of terminating with the corresponding monomer will have its label specifically affected. A scan of the matrix after each step and comparison with the earlier scans will determine which positions correspond to polymers ending with a susceptible monomer. Performing a removal step with a second monomer-specific reagent followed by a scan will identify those positional locations having polymer clusters ending with that second monomer. A similar reagent for the other possible monomers will further define all of the possibilities. Finally, when all of the possible monomers have been removed, the labeling reaction may be repeated and the succession of specific reagent and scanning steps will also be repeated. This procedure allows for a succession of automated steps to determine the sequence of the polymer clusters localized to distinct positions.

Finally, a combination of both specificity of reagent and ability to distinguish label on different monomers can be utilized. Neither alone need be relied upon exclusively. Thus, in the case of nucleotides, an ability to distinguish into two separate classes of nucleotides, e.g., A and C from G and T, combined with specific reagents for distinguishing between the indistinguishable label pairs, e.g., in the

US 7,056,666 B2

21

example provided, A from C, or G from T, can also provide sufficient information for sequencing.

Instead of performing four specific reactions on the same substrate matrix, each of the four individual reactions can be performed on separate parallel matrices. Four separate substrate matrices may be made by a replica plating or successive transfers, each matrix having the same spatial distribution of polymer clusters. Thereby, each separate substrate can be subjected to only a single specific reagent in a highly optimized reaction. On each cycle, one out of the four parallel substrates should show a signal indicating the monomer at the terminal for the cluster at a given matrix position.

Likewise, two parallel substrates can be provided, and each of the parallel substrates is used to determine two of the four possible nucleotides at each position. Instead of treating a single matrix with four separate reactions, this approach allows treating each of two substrates with only two separate reactions. By minimizing the number of reactions to which each chip is exposed, the side reactions will be minimized, the chemistry will be optimized, and the number of cycles through which a matrix will survive will be optimized. This provides an advantage in the number of cycles to which a matrix can be subjected before the signal to noise becomes indistinguishable.

E. Label

The label is important in providing a detectable signal. The signal may be distinguishable among the various monomers by the nature of the signal, e.g., wavelength or other characteristic, as described in Prober et al. (1987) *Science* 238:336-311. A monomer-specific reagent can allow determination of whether each position has a particular terminal monomer by the presence or loss of label.

The label on the monomer may be attached by a noncovalent attachment, but will be preferably attached by a direct covalent attachment. The label will typically be one which is capable of high positional resolution and does not interfere with the nucleotide-specific chemistry or enzymology. Although many different labels may be devised including enzyme linked immunosorbent assays (ELISA), spectrophotometric labels, light producing or other labels, a fluorescent moiety is the preferred form. For example, an avidin/biotin type affinity binding may be useful for attaching a particular label. Alternatively, an antibody may be used which is specific for binding to a particular terminal monomer. A wide variety of other specific reagents can be used to provide a labeling function. See, for example, U.S. Ser. No. 07/624, 114, (a grandchild of which has issued as U.S. Pat. No. 5,800,992) (sequencing by hybridization), which is hereby incorporated herein by reference.

The means of detection utilized will be selected in combination with various other considerations. In some circumstances, a spectroscopic label may be most compatible with a particular monomer. Enzyme linked assays with a spectrophotometric detection system are a workable system. Phosphorescent or light producing assays provide high sensitivity using charged couple devices. Fluorescent systems provide the same advantages, especially where the incident light beam is a laser. The fluorescent label also may provide the added advantage of fluorescing at different wavelengths for the different monomers, providing a convenient means to distinguish between different monomers. Other forms of label may be desired for various reasons, for example, magnetic labels, radioactive labels, heavy metal atoms, optically detectable labels, spectroscopically detectable labels, fluorescent labels, and magnetic labels.

For sequencing nucleic acids by this method, the labeled monomers are simpler than those monomers used for the

22

synthetic method. The blocking group is unnecessary, but terminal specific reagents are more difficult to produce.

The preferred attachment sites will be at the same location as the blocking site, so a combined label and blocking moiety is more preferred. The label will be attached as described, e.g., in U.S. Ser. No., 07/624,114, a grandchild of which has issued as U.S. Pat. No. 5,800,992 (sequencing by hybridization).

Two types of degradation cycles can be used, either non-specific removal of the terminal labeled nucleotide, or a base-specific removal. With the nonspecific removal means, each of the end monomers, when labeled, should be distinguishable from the other three monomer possibilities. This allows for determination of the terminal nucleotide for the cluster localized at a given matrix position. Then the terminal, labeled nucleotides are non-specifically removed and the newly exposed terminal nucleotides will be again distinguishably labeled.

By this scheme, a specific label for each of the different nucleotides may be provided. For example, fluorescent reagents specific for each of the nucleotides may provide a signal with a different wavelength. This will more usually occur when the fluorescent probe is located near the base moiety of the nucleotide. In the scanning step, the regions terminating with each of the four different nucleotides may be determined. Then, a reaction is performed removing the labeled terminal nucleotides from all of the polymers. This removal may be either enzymatic, using a phosphatase, an exonuclease or other similar enzyme, or chemical, using acid, base, or some other, preferably mild, reagent. Again, the reactions are performed which label each of the terminal nucleotides and a scan step repeated in the same manner.

In the base-specific removal scheme, nucleotide-specific removal can be performed. For example, an enzyme which will function to remove only a single modified nucleotide, e.g., a 5'-fluorescein-dAMP-specific exonuclease, is constructed. This may be achieved by proper construction of a catalytic antibody. Other similar reagents may be generated for each of the other labeled nucleotide monomers.

Catalytic or derivatized antibodies to catalyze the removal of the 3'-end or 5'-most fluorescent base in a base-specific manner may be constructed as follows. A recombinant antibody library or a series of monoclonal antibodies is screened with fluorescent donor-quencher substrates. These substrates consist of a fluorescent labeled base (A, C, G, or T) on the 5' or 3' end joined by a 5' to 3' phosphodiester linkage to a second base. A collection of all four possible second bases for each of the four end bases gives the best selection target for the required non-specificity with respect to the second base. The second base is then tethered to an acceptor group in sufficient proximity to quench the fluorescence of the end group. In the presence of a catalytic antibody with cleaving activity, a fluorescent signal occurs from the separation of the quenching group from the terminal fluorescent label. To assure both base and end specificity, the positive monoclonal antibody clones are rescreened against the other substrates.

Upon selection of an antibody exhibiting the desired specificity (or lack thereof), the reactive group for cleavage may be attached. This cleavage reagent may be chemical or enzymatic and will be attached by an appropriate length linker to the antibody binding site in an orientation which is consistent with the steric requirements of both binding and specific cleavage.

Particularly useful specific reagents may be produced by making antibodies specific for each of the four different modified terminal nucleotide bases. These antibodies would

US 7,056,666 B2

23

then specifically bind only to polymers terminating in the appropriate base analog. By combining a cleavage reagent to the specific antibody, a terminal nucleotide specific cleavage reagent is generated.

In one example of the degradative embodiment, all of the polymers may be uniformly labeled at a particular end. Thereafter, a specific removal reaction which removes only a particular nucleotide may be performed, leaving the three other nucleotides labeled. Thereafter, a scanning step is performed through which all regions which had incorporated that particular nucleotide will have lost the label through specific removal. Then, the second specific reagent will be applied which specifically removes the second labeled nucleotide, and the scanning step following that reaction will allow determination of all regions which lose the second particular nucleotide. This process is repeated with reagents specific for each of the last two remaining labeled nucleotides interspersed with scanning steps, thereby providing information on regions with each of the nucleotides located there. Then, the entire process may be repeated by labeling the next terminal nucleotides uniformly. As mentioned below, replication techniques may allow for making four separate but identical matrix substrates. Each substrate may be subjected to single nucleotide-specific reactions, and the scan results correlated with each of the other parallel substrates.

In the degradation scheme, the polynucleotide linkage to the matrix must be more carefully selected such that the free end of the oligonucleotide segments used for attachment will not interfere with the determinations of the target sequence terminus.

F. Utility

The present sequencing method is useful to monitor and check the accuracy and reliability of the synthetic processes described in U.S. Ser. No. 07/362,901 (VLSIPS® parent) and U.S. Ser. No. 07/492,462 (VLSIPS® CIP) (now U.S. Pat. No. 5,143,854). The present method can be used to check the final products synthesized therein, or to label each monomers as they are added stepwise to monitor the efficiency and accuracy of those synthetic methods.

The present invention can also be used to monitor or sequence matrix bound clusters of positionally distinct polymers. This sequencing process provides the capability of simultaneously sequencing a large plurality of distinct polymers which are positionally segregated.

The method will be used to sequence extremely large stretches of polymer, e.g., nucleic acids. A large number of shorter segments of a large sequence can be sequenced with alignment of overlaps either randomly generated, or in an ordered fashion, or particular sequenceable segments of a large segment can be generated. In one approach, a large segment is subcloned into smaller segments and a sufficient number of the randomly generated subclones are sequenced as described herein to provide sequence overlap and ordering of fragments.

In an alternative approach, a large segment can be successively digested to generate a succession of smaller sized subclones with ends separated by defined numbers of monomers. The subclones can be size sorted by a standard separation procedure and the individual samples from a separation device manually or automatically linked to a matrix in a defined positional map. Fractions resulting from size separation can be spatially attached at defined positions, often at adjacent positions. Then polymer sequences at adjacent positions on the matrix will also be known to have ends which differ by, e.g., approximately 25 or 50 or more

24

monomers, thereby providing significantly greater confidence in overlapping sequence data.

III. Specific Embodiments

A specific series of reactions for sequencing a matrix of polynucleotides is described.

A. Synthetic Method

This method involves annealing a primer (common to all the attached sequences by virtue of the cloning construction) near to the 3' end of the unknown target sequences. DNA polymerase, or a similar polymerase, is used to extend the chains by one base by incubation in the presence of dNTP analogs which function as both chain terminators and fluorescent labels. This is done in a one-step process where each of the four dNTP analogs is identified by a distinct dye, such as described in Prober et al. *Science* 238:336-341, or in four steps, each time adding one of the four bases, interspersed with a scanning identification step. When each cluster incorporates the proper one of the four bases and the fluorescence scanning is complete, the matrix is stripped of the label and the chain terminators are deblocked for a next round of base addition. Because the base addition is directed by the template strand, the complementary sequence of the fragments at each address of the matrix is deduced.

(1) Attachment to a Surface.

Both degradative and synthetic sequencing methods begin by obtaining and immobilizing the target fragments of unknown sequence to be determined at specific locations on the surface.

There are several strategies for photo-directed attachment of the DNA strands to the surface in an orientation appropriate for sequencing. A caged biotin technique, see, e.g., U.S. Ser. No. 07/435,316 (caged biotin parent); and U.S. Ser. No. 07/612,671 (caged biotin CIP), is available. Another technique that is especially applicable for the enzymatic synthesis method is to chemically attach a synthetic oligomer by the 5' end to the entire surface (see FIG. 6), to activate it for photocrosslinking (with psoralen, for example) and to anneal the complementary strands and photocrosslink the target strand of unknown sequence (complementary to this oligonucleotide at the 3' end) at the specific location addressed by light. In this case, the oligonucleotide serves as both the attachment linker and as the synthetic primer. A third method is to physically transfer individual nucleic acid samples to selected positions on the matrix, either manually or automatically.

Many sequences in each step are attached by cloning the library into a series of vectors identical except for the sequences flanking the insert. These primers—can be added at the point of amplification of the cloned DNA with chimeric primers.

Alternatively, sequences are attached to a matrix substrate by colony or phage immobilization. This directly transfers the positional distribution on a petri plate to a usable substrate. Colonies representing a shotgun collection of sequences (enough to assure nearly complete coverage by overlap) are spread over (or in) a nutrient surface at a density to give about 100 or more colonies or plaques in several square centimeters, and the colonies are allowed to grow to about 0.1 mm in diameter (the maximum possible density of clusters at this size is ~10,000 colonies/cm²). As described above, replica platings or successive transfers may allow for preparation of multiple matrices with identical positional distributions of polymers. Each separate matrix may then be dedicated to the reactions applicable to a single monomer.

For example, in the use of a phage library, on a petri dish, the transfer substrate surface is treated to release DNA from

US 7,056,666 B2

25

the phage. This is done, e.g., with CHCl_3 vapor, SDS—NaOH, or by heating. Prior to release of DNA, the phage particles are often adsorbed to the surface by way of an antibody to the coat protein that has been immobilized on the surface. This strategy prevents diffusion of the phage from the colonies. The matrix surface is prepared by coating with an oligonucleotide, immobilized to the surface by one end that has homology with the phage vector DNA adjacent to the cloning site.

The matrix surface is juxtaposed to the growth surface, and the phage DNA is allowed to anneal to the immobilized oligonucleotide. The growth surface is removed, and the hybrid is stabilized by psoralen or an equivalent crosslinking reagent.

This method provides an efficient one-step method of placing many DNA fragments onto the detection surface in preparation for sequencing. Although the colonies are not placed in predefined locations, the random arrangement of the clusters allows the final sequence to be assembled from correlation of overlap sequence data derived from sequence data derived from each of the defined positions of each target cluster.

Sequences are, in other embodiments, attached by a manual or automated transfer technique. A few cells from each colony in a library is toothpicked into microliter wells. The plate is heated to $\sim 100^\circ\text{C}$. for a short period to lyse the cells and release the DNA. The plate is cooled and reagents for cyclic amplification of the DNA using, e.g., PCR technology, are added, including primers common to all the cloned sequences. See, e.g., Innis et al. (1990) *PCR Protocols: A Guide to Methods and Applications*, Academic Press, which is hereby incorporated herein by reference. The DNA is amplified asymmetrically by unbalanced primer concentration to yield an excess of one strand for sequencing and attached to a substrate by manual or automated means.

An alternative form of automated localization is described above in positioning of a succession of smaller sized polymers which are manually or automatically linked to the substrate in a pattern reflecting sequence overlaps.

(2) Enzymatic Polymerization Method.

The nucleic acid template is, in some embodiments, attached to the surface by either the 5' or the 3' end, usually by a method as described above. A preferred method of attachment is to anneal the template to an oligonucleotide attached to the surface and to crosslink the template to the oligonucleotide. Oligonucleotide primers are usually synthesized chemically. In this case, the immobilized oligonucleotide may also serve as a primer for polymerization. Because polymerization proceeds 5' to 3' on the primer, the template will be attached by its 3' end, or a site 3' proximal to the region to be sequenced, for the purposes of the description to follow.

Step 1: A DNA-dependent, DNA polymerase such as those used for conventional DNA sequencing, for example, Klenow fragment of *E. coli* DNA Pol, Sequenase (modified T7 DNA polymerase), Taq (*Thermus aquaticus*) DNA polymerase, Bst (*Bacillus stearothermophilus*), DNA polymerase, reverse transcriptase (from AMV, MMLV, RSV, etc.) or other DNA polymerases, and the reaction components appropriate to the particular DNA polymerase selected, are placed in the incubation chamber in direct contact with the surface.

Step 2: Fluorescent chain terminators (analogues of dATP, dCTP, dGTP, and TP, each labeled with fluorophore preferably emitting at a distinguishable wavelength) are added to the reaction at a sufficient concentration and under suitable reaction conditions (time, temperature, pH, ionic species,

26

etc., see Sambrook et al. (1989) *Molecular Cloning*, vols. 1–3, and Prober et al.) to cause essentially all of the chains on the surface to be extended by one base and thereby terminated. Detection of the specific label thereby incorporated into each chain identifies the last base added at each positional address in the matrix.

Step 3: The chain termination should be reversible by some means, such as treatment with light, heat, pH, certain other chemical or biological (enzymatic) reagents, or some combination of these. Typically the chain termination results from a blocking moiety which is labile to mild treatment. By one of these means, the blocked 3'OH of the terminating base must be made available for chain extension in the next round of polymerization.

Step 4: There are several suitable labeled, terminator structures as follows:

- (a) The fluorophore itself functions as the chain terminator by placement on the 3' hydroxyl through a linkage that is easily and efficiently cleaved (removing the label and leaving the free 3'OH) by light, heat, pH shift, etc. The surface is scanned with a scanning system, e.g., the fluorescence detection system described in U.S. Ser. No. 07/492,462 (VLSIPS® CIP) (now U.S. Pat. No. 5,143,854); and U.S. Ser. No. 07/624,120, a divisional of which has issued as U.S. Pat. No. 5,744,101 (automated VLSIPS®). Then, preferably in a single step, the fluorophore is removed and the chain is activated for the next round of base addition.
- (b) The fluorophore is placed in a position other than the 3'OH of the nucleoside, and a different group is placed on the 3'OH of the dNTPs to function as a chain terminator. The fluorophore and the 3' blocking group are removed by the same treatment in a single step (preferably), or they may be removed in separate steps.
- (c) An alternative polymer stepwise synthetic strategy can be employed. In this embodiment, the fluorophores need not be removable and may be attached to irreversible chain terminators. Examples of such compounds for use in sequencing DNA include, but are not limited to, dideoxynucleotide triphosphate analogs as described by Prober et al. (1987) *Science* 238:336–341. A second, unlabeled and reversible, set of terminators is also required. Examples of these compounds are deoxynucleotide triphosphates with small blocking groups such as acetyl, tBOC, NBOC and NVOC on the 3'OH. These groups are easily and efficiently removed under conditions of high or low pH, exposure to light or heat, etc. After each round of base addition and detection, the fluorophores are deactivated by exposure to light under suitable conditions (these chains have their labeling moiety destroyed and remain terminated, taking part in no further reactions). The unlabeled, reversible terminators are unblocked at the 3'OH by the appropriate treatment to allow chain extension in subsequent rounds of elongation. The proportion of chains labeled in each round can be controlled by the concentration ratio of fluorescent to non-fluorescent terminators, and the reaction can be driven to completion with high concentrations of the unlabeled terminators.
- (d) A single dye strategy is used where all the base analog terminators carry the same fluorophore and each is added one at a time: A, C, G, T. The addition of each base is followed by scanning detection and labeled. After all four fluorophore are added, reversal of the termination is performed, allowing for the addition of the next base analog. Then, each scanning step deter-

US 7,056,666 B2

27

mines whether the immediately preceding labeled nucleotide had been incorporated at each distinct position.

The structures of the fluorescent labeled and reversible terminator base analogs are selected to be compatible with efficient incorporation into the growing chains by the particular DNA polymerase(s) chosen to catalyze extension. For example, where two different chain terminators are used, they may be utilized by two different polymerases that are both present during the chain extension step.

Step 5: An optional step is the permanent capping of chain extension failures with high concentrations of dideoxynucleotide triphosphates. This step serves to reduce the background of fluorescence caused by addition of an incorrect base because of inefficient chain extension (termination) at an earlier step.

Step 6: After scanning to determine fluorescence, the fluorophore is removed or deactivated. Deactivation of the fluorophore can be achieved by a photodestruction event. The chain elongation block is reversed (usually by removing a blocking group to expose the 3'OH) by suitable methods that depend on the particular base analogs chosen; and the substrate is washed in preparation for the next round of polymerization.

Step 7: Repeat the cycle.

B. Chain Degradation Method

This method involves labeling the last base of the chain (distal to the surface attachment) with a fluorescent tag followed by base-specific removal. All the polynucleotide clusters on the matrix are labeled using a standard labeling moiety. Base-specific removal of the last base of each chain, interspersed with fluorescence scanning of the array, will reveal the disappearance of fluorescence and hence the identity of the last base of each chain. When all four labeled end bases have been removed, the polymers attached to the matrix are relabeled and the process is repeated, working successively on the DNA chains.

Alternatively, if the label allows distinguishing between different monomers, simpler degradation processes may be employed. A single scan step can distinguish between all four possible terminal nucleotides. The four separate removal steps are then combined into a single nonspecific terminal nucleotide removal step.

The DNA will usually be attached to the substrate by the 3' or 5' terminus depending on the scheme of labeling and cleavage. Because there are well-known 5'-labeling methods, see, e.g., Gait (1984) *Oligonucleotide Synthesis: A Practical Approach*, IRL Press, Oxford, this discussion will assume the 3' end is attached to the substrate with the 5' end free.

Step 1: All the 5'-end bases are labeled with 5'-specific chemistry, e.g., 5' amino linkage to FITC, Nelson et al. (1989) *Nucl. Acids Res.* 17:7179-7186, which is hereby incorporated herein by reference.

Step 2: Scan the matrix to obtain the background level.

Step 3: Optional: Cap all of the labeling failures, e.g., polymers whose ends were not labeled.

Step 4: The terminal A's are removed with end-base, A-specific reagents (such a reagent may be chemical or biological). One example is a 5'-fluorescein-dAMP-specific exonuclease made as a catalytic antibody (see the description above for a scheme of producing this reagent).

Step 5: Scan the matrix to detect those chains that had terminated in A (these will be reduced in fluorescence compared to the fluorescent labeled background).

Step 6: Repeat steps 4 and 5 for each of other three possible bases using the appropriate fluorescein-base-spe-

28

cific cleavage reagent and scan after removal of each of the C's, the G's, and the T's. This succession of steps will allow the determination of the terminal nucleotide of each positionally defined cluster.

Step 7: Relabel the 5' terminal nucleotide of all the new end bases that have been exposed by the earlier rounds of cleavage, and repeat the stepwise removal and scanning processes.

This approach can be extended to protein sequencing using 20 catalytic antibodies (or other amino acid-specific cleavage reagents), each recognizing a terminal amino acid and removing that terminal residue.

The process for sequencing may be summarized as follows for enzymatic polymerization:

1) Target DNA templates (to be sequenced) are attached at positionally defined locations on the matrix substrate.

2) Fluorescent chain terminators are added to a primer under conditions where all polymer chains are terminated after addition of the next base complementary to the template.

3) The matrix is scanned to determine which base was added to each location. This step correlates the added base with a position on the matrix.

4) Chains failing to extend (and therefore to terminate) are capped.

5) The fluorophores are removed or deactivated.

6) The terminators are activated for further chain extension, usually by removal of a blocking group.

7) Steps 2 through 6 are repeated to obtain the base-by-base sequence of many different positionally separated DNA fragments simultaneously.

C. Screening for New Nucleotide Analog/Polymerase Combinations.

The use of a functional combination of blocked nucleotide with a polymerase is important in the synthetic embodiment of the present invention. It is important to ensure that only a single nucleotide is incorporated at the appropriate step. The following protocol describes how to screen for a functional combination.

Test 1. (Test for Polymerase Inhibition)

In a reaction volume of 20 μ l, mix

1 μ g M13mp19 single stranded DNA template

2.5 ng standard M13 primer (17-mer:

5'-GTTTTCCAGTCACGAC-3' (SEQ ID NO:1)

60 mM tris-CL pH 8.5

7.5 mM $MgCl_2$

75 mM NaCl

Template and primer are annealed by heating to 95° C., then cooling to -25°

Extension components are added:

50 μ M (each) dATP, dCTP, dGTP, TTP

10 μ Ci P32 dATP

0.01 μ M to 1 mM of the putative terminator compound, further titrations may be desired.

20 units AMV reverse transcriptase water to 20 μ l final volume

The reaction is run at 42° for about 30 minutes.

Aliquots are taken at 10, 20, 30 minutes, and samples are TCA precipitated after the addition of 10 μ g tRNA carrier.

The filters are counted for acid-precipitable radioactivity and the mass of dATP incorporated is calculated as a function of reaction time.

Control reactions are run in parallel consisting of

A) no added terminator

B) 10 μ M and 100 μ M

US 7,056,666 B2

29

The termination activity of the experimental samples relative to that of ddNTPs is estimated, and a nucleotide is appropriate for further testing if it substantially decreases the number of acid precipitable counts at any time or relative concentration.

Test 2 (Test for Base Specific Termination Activity)

Reactions are run essentially as described by Prober et al. except:

1. Unlabelled primer is used
2. 1 μ Ci P32 dATP is included
3. No dideoxynTPs are added to the experimental samples (control reactions containing ddNTP at the usual concentrations, and no test terminators are run in parallel)
4. The test compound is added at a concentration estimated to give 1% and 10% inhibition of incorporation as determined by test #1.

The reactions are run for 10 min at 42°. 100 μ M dNTPS are added and the reaction run for an additional 10 min. A portion of the reaction is prepared and run on a sequencing gel in the usual fashion. The ladders obtained with the test compound are compared with that obtained in the ddNTP reactions and the fidelity of the termination activity of the test compound is thereby assessed.

IV. Apparatus

The present invention provides a new use for an apparatus comprising a reaction chamber and a scanning apparatus which can scan a substrate material exposed to the chamber. FIG. 11 illustrates a system and a schematized reaction chamber to which is attached a silicon or glass substrate. The system has a detection system 102 as illustrated, in one embodiment, in FIG. 7. A silicon substrate 104, is attached against and forming a seal to make a reaction chamber 106. Leading into and out of the chamber are tubes 108, with valves 110 which control the entry and exit of reagents 112 which are involved in the stepwise reactions. The chamber is held at a constant temperature by a temperature block 114.

All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference. The invention now being fully described, it will be apparent to one of ordinary skill in the art that many changes and modifications can be made thereto without departing from the spirit or scope of the claims.

30

What is claimed is:

1. A method for determining the positions of polynucleotides which terminate with a given monomer, comprising providing a plurality of different polynucleotides attached at distinct positions of a surface of a support by annealing of the polynucleotides to oligonucleotides immobilized to regions of the surface of the support at a density of at least 1000 regions/cm², the polynucleotides having terminal monomers labeled in a monomer-type specific manner; and scanning said surface, thereby determining the positions of said labeled monomer.

2. A method for concurrently determining which subset of a plurality of different polymers attached at distinct positions of a surface of a solid substrate at separable locations at a density of at least 1000 locations per cm² terminates with a given terminal subunit, said method comprising the steps of:

mixing said solid substrate with a solution comprising a reagent, and thereby selectively marking positionally distinct polymers which terminate with said given terminal subunit; and

determining with a detector which separable locations are marked, thereby determining which subset of said positionally distinct polymers terminated with said given terminal subunit.

3. A method of claim 2, wherein said solution comprises a reagent which marks said positionally distinct polymer with a fluorescent label moiety.

4. A method of claim 2, wherein said terminal subunit is selected from the group consisting of adenosine, cytosine, guanosine, and thymine.

5. A method for determining which subset of a plurality of primer polynucleotides have a predetermined oligonucleotide, wherein the polynucleotides are complementary to and annealed to different template strands which are attached at distinct positions of a surface of a solid substrate at a density of at least 1000 locations per cm², said method comprising the steps of:

selectively marking said subset of primer polynucleotides having the predetermined oligonucleotide; and detecting which polynucleotides are marked.

6. A method of claim 5, wherein said oligonucleotide subunit is a single nucleotide.

7. A method of claim 6, wherein the marking comprises elongating said primer with a labeled nucleotide which is complementary to a template.

8. A method of claim 5, wherein the marking step uses a polymerase and a blocked and labeled adenine.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 1

<210> SEQ ID NO 1

<211> LENGTH: 17

<212> TYPE: DNA

<213> ORGANISM: Artificial sequence

<220> FEATURE:

<223> OTHER INFORMATION: M13 primer

<400> SEQUENCE: 1

gttttcccg tcacgac

US 7,056,666 B2

31

9. The method of claim 1, wherein the labeled terminal monomers can be distinguished from one another by the wavelength of fluorescent emission.

10. The method of claim 1, wherein the support is a material selected from the group consisting of silicon, glass, ceramics and plastics.

11. A method for determining a subset of polynucleotides which terminate with a given monomer, comprising

providing a plurality of different polynucleotides attached at different positions of a surface of a support by annealing of the polynucleotides to oligonucleotides immobilized to regions of the surface at a density of at least 1000 regions per cm², the polynucleotides having terminal monomers labeled:

32

in a monomer type specific manner; and scanning said surface, thereby determining the subset of polynucleotides terminate with a given monomer.

12. The method of claim 11, wherein the labeled terminal monomers can be distinguished from one another by the wavelength of fluorescent emission.

13. The method of claim 11, wherein the support is a material selected from the group consisting of silicon, glass, ceramics and plastics.

14. The method of claim 1, wherein the oligonucleotides are synthesized oligonucleotides.

15. The method of claim 11, wherein the oligonucleotides are synthesized oligonucleotides.

* * * * *

JS 44 (Rev. 11/04)

CIVIL COVER SHEET

The JS 44 civil cover sheet and the information contained herein neither replace nor supplement the filing and service of pleadings or other papers as required by law, except as provided by local rules of court. This form, approved by the Judicial Conference of the United States in September 1974, is required for the use of the Clerk of Court for the purpose of initiating the civil docket sheet. (SEE INSTRUCTIONS ON THE REVERSE OF THE FORM.)

I. (a) PLAINTIFFS Affymetrix, Inc.	DEFENDANTS Illumina, Inc.
(b) County of Residence of First Listed Plaintiff _____ (EXCEPT IN U.S. PLAINTIFF CASES)	County of Residence of First Listed Defendant _____ (IN U.S. PLAINTIFF CASES ONLY) NOTE: IN LAND CONDEMNATION CASES, USE THE LOCATION OF THE LAND INVOLVED.
(c) Attorney's (Firm Name, Address, and Telephone Number) Maryellen Noreika, MORRIS, NICHOLS, ARSHT & TUNNELL LLP, 1201 North Market Street, P.O. Box 1347, Wilmington, DE 19899-1347, (302) 658-9200	Attorneys (If Known)

II. BASIS OF JURISDICTION (Place an "X" in One Box Only)	III. CITIZENSHIP OF PRINCIPAL PARTIES (Place an "X" in One Box for Plaintiff and One Box for Defendant)																								
<input type="checkbox"/> 1 U.S. Government Plaintiff <input type="checkbox"/> 2 U.S. Government Defendant <input checked="" type="checkbox"/> 3 Federal Question (U.S. Government Not a Party) <input type="checkbox"/> 4 Diversity (Indicate Citizenship of Parties in Item III)	<table style="width: 100%;"> <tr> <th></th> <th>PTF</th> <th>DEF</th> <th></th> <th>PTF</th> <th>DEF</th> </tr> <tr> <td>Citizen of This State</td> <td><input type="checkbox"/> 1</td> <td><input type="checkbox"/> 1</td> <td>Incorporated or Principal Place of Business In This State</td> <td><input type="checkbox"/> 4</td> <td><input type="checkbox"/> 4</td> </tr> <tr> <td>Citizen of Another State</td> <td><input type="checkbox"/> 2</td> <td><input type="checkbox"/> 2</td> <td>Incorporated and Principal Place of Business In Another State</td> <td><input type="checkbox"/> 5</td> <td><input type="checkbox"/> 5</td> </tr> <tr> <td>Citizen or Subject of a Foreign Country</td> <td><input type="checkbox"/> 3</td> <td><input type="checkbox"/> 3</td> <td>Foreign Nation</td> <td><input type="checkbox"/> 6</td> <td><input type="checkbox"/> 6</td> </tr> </table>		PTF	DEF		PTF	DEF	Citizen of This State	<input type="checkbox"/> 1	<input type="checkbox"/> 1	Incorporated or Principal Place of Business In This State	<input type="checkbox"/> 4	<input type="checkbox"/> 4	Citizen of Another State	<input type="checkbox"/> 2	<input type="checkbox"/> 2	Incorporated and Principal Place of Business In Another State	<input type="checkbox"/> 5	<input type="checkbox"/> 5	Citizen or Subject of a Foreign Country	<input type="checkbox"/> 3	<input type="checkbox"/> 3	Foreign Nation	<input type="checkbox"/> 6	<input type="checkbox"/> 6
	PTF	DEF		PTF	DEF																				
Citizen of This State	<input type="checkbox"/> 1	<input type="checkbox"/> 1	Incorporated or Principal Place of Business In This State	<input type="checkbox"/> 4	<input type="checkbox"/> 4																				
Citizen of Another State	<input type="checkbox"/> 2	<input type="checkbox"/> 2	Incorporated and Principal Place of Business In Another State	<input type="checkbox"/> 5	<input type="checkbox"/> 5																				
Citizen or Subject of a Foreign Country	<input type="checkbox"/> 3	<input type="checkbox"/> 3	Foreign Nation	<input type="checkbox"/> 6	<input type="checkbox"/> 6																				

IV. NATURE OF SUIT (Place an "X" in One Box Only)							
CONTRACT <input type="checkbox"/> 110 Insurance <input type="checkbox"/> 120 Marine <input type="checkbox"/> 130 Miller Act <input type="checkbox"/> 140 Negotiable Instrument <input type="checkbox"/> 150 Recovery of Overpayment & Enforcement of Judgment <input type="checkbox"/> 151 Medicare Act <input type="checkbox"/> 152 Recovery of Defaulted Student Loans (Excl. Veterans) <input type="checkbox"/> 153 Recovery of Overpayment of Veteran's Benefits <input type="checkbox"/> 160 Stockholders' Suits <input type="checkbox"/> 190 Other Contract <input type="checkbox"/> 195 Contract Product Liability <input type="checkbox"/> 196 Franchise	TORTS <table style="width: 100%;"> <tr> <td style="vertical-align: top;"> PERSONAL INJURY <input type="checkbox"/> 310 Airplane <input type="checkbox"/> 315 Airplane Product Liability <input type="checkbox"/> 320 Assault, Libel & Slander <input type="checkbox"/> 330 Federal Employers' Liability <input type="checkbox"/> 340 Marine <input type="checkbox"/> 345 Marine Product Liability <input type="checkbox"/> 350 Motor Vehicle <input type="checkbox"/> 355 Motor Vehicle Product Liability <input type="checkbox"/> 360 Other Personal Injury </td> <td style="vertical-align: top;"> PERSONAL INJURY <input type="checkbox"/> 362 Personal Injury - Med. Malpractice <input type="checkbox"/> 365 Personal Injury - Product Liability <input type="checkbox"/> 368 Asbestos Personal Injury Product Liability PERSONAL PROPERTY <input type="checkbox"/> 370 Other Fraud <input type="checkbox"/> 371 Truth in Lending <input type="checkbox"/> 380 Other Personal Property Damage <input type="checkbox"/> 385 Property Damage Product Liability </td> </tr> </table>	PERSONAL INJURY <input type="checkbox"/> 310 Airplane <input type="checkbox"/> 315 Airplane Product Liability <input type="checkbox"/> 320 Assault, Libel & Slander <input type="checkbox"/> 330 Federal Employers' Liability <input type="checkbox"/> 340 Marine <input type="checkbox"/> 345 Marine Product Liability <input type="checkbox"/> 350 Motor Vehicle <input type="checkbox"/> 355 Motor Vehicle Product Liability <input type="checkbox"/> 360 Other Personal Injury	PERSONAL INJURY <input type="checkbox"/> 362 Personal Injury - Med. Malpractice <input type="checkbox"/> 365 Personal Injury - Product Liability <input type="checkbox"/> 368 Asbestos Personal Injury Product Liability PERSONAL PROPERTY <input type="checkbox"/> 370 Other Fraud <input type="checkbox"/> 371 Truth in Lending <input type="checkbox"/> 380 Other Personal Property Damage <input type="checkbox"/> 385 Property Damage Product Liability	FORFEITURE/PENALTY <input type="checkbox"/> 610 Agriculture <input type="checkbox"/> 620 Other Food & Drug <input type="checkbox"/> 625 Drug Related Seizure of Property 21 USC 881 <input type="checkbox"/> 630 Liquor Laws <input type="checkbox"/> 640 R.R. & Truck <input type="checkbox"/> 650 Airline Regs. <input type="checkbox"/> 660 Occupational Safety/Health <input type="checkbox"/> 690 Other	BANKRUPTCY <input type="checkbox"/> 422 Appeal 28 USC 158 <input type="checkbox"/> 423 Withdrawal 28 USC 157 PROPERTY RIGHTS <input type="checkbox"/> 820 Copyrights <input checked="" type="checkbox"/> 830 Patent <input type="checkbox"/> 840 Trademark SOCIAL SECURITY <input type="checkbox"/> 861 HIA (1395ff) <input type="checkbox"/> 862 Black Lung (923) <input type="checkbox"/> 863 DIWC/DIWW (405(g)) <input type="checkbox"/> 864 SSID Title XVI <input type="checkbox"/> 865 RSI (405(g))	OTHER STATUTES <input type="checkbox"/> 400 State Reapportionment <input type="checkbox"/> 410 Antitrust <input type="checkbox"/> 430 Banks and Banking <input type="checkbox"/> 450 Commerce <input type="checkbox"/> 460 Deportation <input type="checkbox"/> 470 Racketeer Influenced and Corrupt Organizations <input type="checkbox"/> 480 Consumer Credit <input type="checkbox"/> 490 Cable/Sat TV <input type="checkbox"/> 810 Selective Service <input type="checkbox"/> 850 Securities/Commodities/Exchange <input type="checkbox"/> 875 Customer Challenge 12 USC 3410 <input type="checkbox"/> 890 Other Statutory Actions <input type="checkbox"/> 891 Agricultural Acts <input type="checkbox"/> 892 Economic Stabilization Act <input type="checkbox"/> 893 Environmental Matters <input type="checkbox"/> 894 Energy Allocation Act <input type="checkbox"/> 895 Freedom of Information Act <input type="checkbox"/> 900 Appeal of Fee Determination Under Equal Access to Justice <input type="checkbox"/> 950 Constitutionality of State Statutes	REAL PROPERTY <input type="checkbox"/> 210 Land Condemnation <input type="checkbox"/> 220 Foreclosure <input type="checkbox"/> 230 Rent Lease & Ejectment <input type="checkbox"/> 240 Torts to Land <input type="checkbox"/> 245 Tort Product Liability <input type="checkbox"/> 290 All Other Real Property
PERSONAL INJURY <input type="checkbox"/> 310 Airplane <input type="checkbox"/> 315 Airplane Product Liability <input type="checkbox"/> 320 Assault, Libel & Slander <input type="checkbox"/> 330 Federal Employers' Liability <input type="checkbox"/> 340 Marine <input type="checkbox"/> 345 Marine Product Liability <input type="checkbox"/> 350 Motor Vehicle <input type="checkbox"/> 355 Motor Vehicle Product Liability <input type="checkbox"/> 360 Other Personal Injury	PERSONAL INJURY <input type="checkbox"/> 362 Personal Injury - Med. Malpractice <input type="checkbox"/> 365 Personal Injury - Product Liability <input type="checkbox"/> 368 Asbestos Personal Injury Product Liability PERSONAL PROPERTY <input type="checkbox"/> 370 Other Fraud <input type="checkbox"/> 371 Truth in Lending <input type="checkbox"/> 380 Other Personal Property Damage <input type="checkbox"/> 385 Property Damage Product Liability						
CIVIL RIGHTS <input type="checkbox"/> 441 Voting <input type="checkbox"/> 442 Employment <input type="checkbox"/> 443 Housing/Accommodations <input type="checkbox"/> 444 Welfare <input type="checkbox"/> 445 Amer. w/Disabilities - Employment <input type="checkbox"/> 446 Amer. w/Disabilities - Other <input type="checkbox"/> 440 Other Civil Rights	PRISONER PETITIONS <input type="checkbox"/> 510 Motions to Vacate Sentence Habeas Corpus: <input type="checkbox"/> 530 General <input type="checkbox"/> 535 Death Penalty <input type="checkbox"/> 540 Mandamus & Other <input type="checkbox"/> 550 Civil Rights <input type="checkbox"/> 555 Prison Condition	LABOR <input type="checkbox"/> 710 Fair Labor Standards Act <input type="checkbox"/> 720 Labor/Mgmt. Relations <input type="checkbox"/> 730 Labor/Mgmt. Reporting & Disclosure Act <input type="checkbox"/> 740 Railway Labor Act <input type="checkbox"/> 790 Other Labor Litigation <input type="checkbox"/> 791 Empl. Ret. Inc. Security Act	FEDERAL TAX SUITS <input type="checkbox"/> 870 Taxes (U.S. Plaintiff or Defendant) <input type="checkbox"/> 871 IRS—Third Party 26 USC 7609				

V. ORIGIN (Place an "X" in One Box Only)						
<input checked="" type="checkbox"/> 1 Original Proceeding	<input type="checkbox"/> 2 Removed from State Court	<input type="checkbox"/> 3 Remanded from Appellate Court	<input type="checkbox"/> 4 Reinstated or Reopened	<input type="checkbox"/> 5 Transferred from another district (specify)	<input type="checkbox"/> 6 Multidistrict Litigation	<input type="checkbox"/> 7 Appeal to District Judge from Magistrate Judgment

VI. CAUSE OF ACTION	Cite the U.S. Civil Statute under which you are filing (Do not cite jurisdictional statutes unless diversity): 35 U.S.C. Section 271 Brief description of cause: patent infringement
----------------------------	---

VII. REQUESTED IN COMPLAINT:	<input type="checkbox"/> CHECK IF THIS IS A CLASS ACTION UNDER F.R.C.P. 23	DEMAND \$	CHECK YES only if demanded in complaint: JURY DEMAND: <input checked="" type="checkbox"/> Yes <input type="checkbox"/> No
-------------------------------------	--	------------------	---

VIII. RELATED CASE(S) IF ANY	(See instructions): JUDGE <u>Farnan</u>	DOCKET NUMBER <u>04-901</u>
-------------------------------------	---	-----------------------------

DATE

SIGNATURE OF ATTORNEY OF RECORD

10/24/07

Maryellen Noreika

FOR OFFICE USE ONLY

RECEIPT # _____ AMOUNT _____ APPLYING IFP _____ JUDGE _____ MAG. JUDGE _____

INSTRUCTIONS FOR ATTORNEYS COMPLETING CIVIL COVER SHEET FORM JS 44**Authority For Civil Cover Sheet**

The JS 44 civil cover sheet and the information contained herein neither replaces nor supplements the filings and service of pleading or other papers as required by law, except as provided by local rules of court. This form, approved by the Judicial Conference of the United States in September 1974, is required for the use of the Clerk of Court for the purpose of initiating the civil docket sheet. Consequently, a civil cover sheet is submitted to the Clerk of Court for each civil complaint filed. The attorney filing a case should complete the form as follows:

I. (a) Plaintiffs-Defendants. Enter names (last, first, middle initial) of plaintiff and defendant. If the plaintiff or defendant is a government agency, use only the full name or standard abbreviations. If the plaintiff or defendant is an official within a government agency, identify first the agency and then the official, giving both name and title.

(b) County of Residence. For each civil case filed, except U.S. plaintiff cases, enter the name of the county where the first listed plaintiff resides at the time of filing. In U.S. plaintiff cases, enter the name of the county in which the first listed defendant resides at the time of filing. (NOTE: In land condemnation cases, the county of residence of the "defendant" is the location of the tract of land involved.)

(c) Attorneys. Enter the firm name, address, telephone number, and attorney of record. If there are several attorneys, list them on an attachment, noting in this section "(see attachment)".

II. Jurisdiction. The basis of jurisdiction is set forth under Rule 8(a), F.R.C.P., which requires that jurisdictions be shown in pleadings. Place an "X" in one of the boxes. If there is more than one basis of jurisdiction, precedence is given in the order shown below.

United States plaintiff. (1) Jurisdiction based on 28 U.S.C. 1345 and 1348. Suits by agencies and officers of the United States are included here.

United States defendant. (2) When the plaintiff is suing the United States, its officers or agencies, place an "X" in this box.

Federal question. (3) This refers to suits under 28 U.S.C. 1331, where jurisdiction arises under the Constitution of the United States, an amendment to the Constitution, an act of Congress or a treaty of the United States. In cases where the U.S. is a party, the U.S. plaintiff or defendant code takes precedence, and box 1 or 2 should be marked.

Diversity of citizenship. (4) This refers to suits under 28 U.S.C. 1332, where parties are citizens of different states. When Box 4 is checked, the citizenship of the different parties must be checked. (See Section III below; federal question actions take precedence over diversity cases.)

III. Residence (citizenship) of Principal Parties. This section of the JS 44 is to be completed if diversity of citizenship was indicated above. Mark this section for each principal party.

IV. Nature of Suit. Place an "X" in the appropriate box. If the nature of suit cannot be determined, be sure the cause of action, in Section VI below, is sufficient to enable the deputy clerk or the statistical clerks in the Administrative Office to determine the nature of suit. If the cause fits more than one nature of suit, select the most definitive.

V. Origin. Place an "X" in one of the seven boxes.

Original Proceedings. (1) Cases which originate in the United States district courts.

Removed from State Court. (2) Proceedings initiated in state courts may be removed to the district courts under Title 28 U.S.C., Section 1441. When the petition for removal is granted, check this box.

Remanded from Appellate Court. (3) Check this box for cases remanded to the district court for further action. Use the date of remand as the filing date.

Reinstated or Reopened. (4) Check this box for cases reinstated or reopened in the district court. Use the reopening date as the filing date.

Transferred from Another District. (5) For cases transferred under Title 28 U.S.C. Section 1404(a). Do not use this for within district transfers or multidistrict litigation transfers.

Multidistrict Litigation. (6) Check this box when a multidistrict case is transferred into the district under authority of Title 28 U.S.C. Section 1407. When this box is checked, do not check (5) above.

Appeal to District Judge from Magistrate Judgment. (7) Check this box for an appeal from a magistrate judge's decision.

VI. Cause of Action. Report the civil statute directly related to the cause of action and give a brief description of the cause. **Do not cite jurisdictional statutes unless diversity.** Example: U.S. Civil Statute: 47 USC 553
Brief Description: Unauthorized reception of cable service

VII. Requested in Complaint. Class Action. Place an "X" in this box if you are filing a class action under Rule 23, F.R.Cv.P.

Demand. In this space enter the dollar amount (in thousands of dollars) being demanded or indicate other demand such as a preliminary injunction.

Jury Demand. Check the appropriate box to indicate whether or not a jury is being demanded.

VIII. Related Cases. This section of the JS 44 is used to reference related pending cases if any. If there are related pending cases, insert the docket numbers and the corresponding judge names for such cases.

Date and Attorney Signature. Date and sign the civil cover sheet.

AO FORM 85 RECEIPT (REV. 9/04)

United States District Court for the District of Delaware

Civil Action No. 07-670-

ACKNOWLEDGMENT
OF RECEIPT FOR AO FORM 85

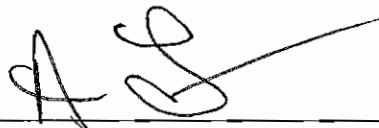
NOTICE OF AVAILABILITY OF A
UNITED STATES MAGISTRATE JUDGE
TO EXERCISE JURISDICTION

FILED
CLERK U.S. DISTRICT COURT
DISTRICT OF DELAWARE
2007 OCT 24 AM 11:15

I HEREBY ACKNOWLEDGE RECEIPT OF 2 COPIES OF AO FORM 85.

OCT 24 2007

(Date forms issued)



(Signature of Party or their Representative)

Aaron Johnston

(Printed name of Party or their Representative)

Note: Completed receipt will be filed in the Civil Action